دانشگاه علوم پزشکی بوشهر

کارگاه مقدماتی SPSS

تحلیل داده های آماری با

SPSS

جزوه درسی

جمع آوری، تهیه و تنظیم: دکتر طهماسبی

چهارم تا پنجم فروردین ۱۳۹۲

مقدمه:

امروزه با توجه به توسعه علوم و گسترش تکنولوژی، استفاده از علم آمار در علوم مختلف بیش از پیش مورد توجه قرار گرفته است و آموختن روشهای تحلیل آماری برای تمامی علوم جزء ملزومات گردیده است، فرآیند آنالیز آماری کمک میکند تا پژوهشگر بتواند از داده های اولیه مربوط به مشاهدات نمونه ای، اطلاعات مورد نیاز خود را استخراج کند و در صورت لزوم نتایج را تعمیم دهد. اگر حجم داده ها زیاد باشد، استفاده از روشهای مختلف تحلیلهای آماری بصورت دستی بسیار خسته کننده و مشکل خواهد بود. امروزه با توجه به وجود انواع مختلف نرم افزارهای آماری مشکل مرتفع گردیده، و ما را قادر به انجام انواع تحلیلهای پیچیده آماری نموده است،

نرم افزار SPSS یکی از قدیمی ترین، برنامه های کاربردی در زمینه تجزیه وتحلیلهای آماری است؛ واژه SPSS مخفف Statistical package for social science (نرم افزار آماری برای علوم اجتماعی) می باشد. این نرم‌افزار که یکی از نرم‌افزارهای تخصصی آمار است، جهت تحلیل داده های آماری در حیطهٔ علوم اجتماعی، روانشناسی و علوم رفتاری و ... مورد استفاده قرار می گیرد.

نرم افزاری آماری با قابلیتهای انجام توصیفی زیبا و گویا از اطلاعات، شامل رسم نمودارها و چارت های گوناگون و محاسبات مربوط به میانگین، انحراف معیار، واریانس، میانه و سایر تحلیلها آماری می باشد.

قابلیت‌های نرم افزار SPSS به شرح زیر است:

- تهیه خلاصه‌های آماری مانند گراف‌ها، جداول، آماره‌ها و ...
- انواع توابع ریاضی مانند قدر مطلق، تابع علامت، لگاریتم، توابع مثلثاتی و...
- تهیه انواع جداول سفارشی مانند جداول فراوانی، فراوانی تجمعی، درصد فراوانی و...
- انواع توزیع‌های آماری شامل توزیع‌های گسسته و پیوسته
- تهیه انواع طرح‌های آماری
- انجام آنالیز واریانس یکطرفه، دوطرفه، چندطرفه و آنالیز کوواریانس
- تکنیک‌های تجزیه و تحلیل سری‌های زمانی
- ایجاد داده‌های تصادفی و پیوسته
- محاسبه انواع آماره‌های توصیفی
- انواع آزمون‌های مرتبط با مقایسه میانگین بین دو یا چند جامعه مستقل و وابسته
- قابلیت مبادله اطلاعات با نرم‌افزارهای دیگر
- برازش انواع مختلف رگرسیون

آمار علم و عمل توسعه دانش انسانی از طریق استفاده از داده های تجربی است. آمار مطالعه لذت بخشی است در باب این موضوع که چگونه می توان جهان ناشناخته ای را با گشودن چند دریچه به روی آن توصیف کرد. با پرداختن به آمار لذت فکر کردن به یک شیوه کاملا جدید را کشف خواهیم کرد. به عبارتی می توان گفت که آمار استنباطی به شیوه‌هایی اطلاق می‌شود که از طریق آنها ویژگیهای گروههای بزرگ بر اساس اندازه گیری همان ویژگیها در گروههای کوچک استنباط می‌شود.

بطور مثال:

چه روش آموزشی برای گروه سنی از دانش آموزان مناسب است؟

توزیع بهره هوشی در یک جامعه چگونه است؟

در پژوهش‌های روان شناسی و سایر علوم رفتاری کسب اطلاعات درباره گروههـای کوچـک غالبـا هـدف پژوهشـگر نیست، بلکه او علاقمند است که از طریق یافته‌های این گروه کوچک، اطلاعات لازم را درباره جامعه‌ای که ایـن گـروه کوچک را از آن انتخاب کرده است، کسب کند. به عبارت دیگر در این پژوهش‌ها هدف پژوهشگر تعمیم نتایج بدسـت آمده از یک گروه کوچک به یک جامعه بزرگتر می‌باشد. این تعمیم مستلزم آن است که پژوهشگر از روش‌های آماری پیشرفته تری تحت عنوان آمار استنباطی (Inferential Statistics) استفاده نماید .

زمانی به استفاده از روشهای آماری رو می آوریم که:

۱) بخواهیم داده ها را به صورت یک مجموعه، خلاصه کرده و توصیف نماییم.
۲) بخواهیم اطمینان دهیم که در شرایطی کـه یـک پـروژه دقیقـاً بـه همـان شـکل تکـرار شـود، همـان یافتـه هـا بدست خواهد آمد.

اهداف اصلی آمار عبارتند از:

۱) انجام استنباط درباره جامعه از طریق تجزیه و تحلیل اطلاعات موجود در داده های نمونه ای.
۲) سنجش میزان عدم حتمیتی که در این استنباط ها وجود دارد. علمی کـه بـرای رسـیدن بـه هـدف هـای فـوق اهمیت دارد, عبارتست از طرح ریزی فرآیند و دامنه نمونه گیری به طوریکه مشاهدات مبنـایی بـرای اسـتخراج استنباط های معتبر تشکیل دهند.

برای انجام استنباط آماری با مفاهیم بسیاری روبه رو هستیم. در زیر به بیان تعـدادی از ایـن مفـاهیم پرداختـه و در ادامه چند آزمون آماری متداول را مورد بررسی قرار می دهیم.

یک مفهوم مهم در این رابطه ***نمونه*** (Sample) می باشد. یک نمونه مجموعه ای از مشاهدات (که اغلب بـه شـکل تصادفی انتخاب می شوند) است که مـی توانـد در یـک مجموعـه مرجـع یـا ***جامعـه*** (Population) از مشـاهدات احتمالی انجام شود. در قیاس با قرعه کشی، می توان یک جامعه را مانند اعداد فرض کرد که در گردونه قرار دارنـد و در این حالت اعدادی که انتخاب می شوند به عنوان نمونه در نظر گرفت. نمونه های گرفته شده از یک جامعه ثابـت، می توانند از حیث ویژگیهایشان بسیار متفاوت باشند. بنابراین یک نمونـه تصـادفی الزامـاً نماینـده واقعـی از جامعـه مرجع نیست و ممکن است ویژگیهای بسیار متفاوتی (مانند میانگین و انحـراف معیـار) ازآن داشـته باشـد. مقـادیر مربوط به ویژگیهای یک نمونه (میانگین، انحراف معیار و غیره) را***آماره*** (Statistics) و ویژگیهای معـادل آنهـا در جامعه اصلی را ***پارامتر*** (Parameters) می نامند. در تحقیقات، مـا همیشـه بـه دنبـال ویژگیهـای جامعـه اصـلی (پارامترها) و نه ویژگیهای نمونه (آماره) هستیم.

آشنایی با روش آمار استنباطی

استنباط آماری که در واقع یک نوع نتیجه گیری کلی از جزء به کـل اسـت، در معـرض آزمـایش و خطاسـت. آمـار استنباطی به منظور برآورد پارامترهای جامعه (مثلاً میانگین جامعه) از طریق **نمونه گیری علمی** از جامعه مورد نظر و همچنین مقایسه پارامترهای دو و یا بیشتر از دو جامعه بکار می‌رود.

برآورد

یک جنبه از استنباط آماری محاسبه برآوردهایی (Estimates) از پارمترهای جامعـه ماننـد میـانگین جامعـه از طریق آماره های نمونه ها مانند میانگین نمونه است. بنابراین اگر از جامعه‌ای نمونه انتخاب کنیم، میانگین این نمونـه

را به منظور برآورد میانگین جامعه در نطر می گیریم، در واقع یک برآورد یا پیش بینی در باره میانگین جامعه از طریق نمونه انتخابی انجام داده‌ایم.
به کمیتی که مبتنی بر آن پارامتر جامعه برآورد می گردد، برآوردگر یا آماره اطلاق می شود. یک برآوردگر مناسب باید دارای ویژگیهایی باشد یعنی برآوردگر مناسب باید بدون سوگیر(Unbiased) ، با ثبات(Consistent) ، کارا (Efficient)و مکفی (Sufficent) باشد.

آزمون فرض

فرضیه آماری نقطه آغاز آزمون فرض است. فرضیه آماری یک بیان مقداری در باره پارامترهای جامعه است و اصولا بدون داشتن فرضیه آماری امکان انجام یک آزمون دشوار است. فرضیه آماری به دو دسته فرض صفر (H_0) و فرض مخالف (H_1) بیان می‌شود. اغلب فرضیه بیانگر این مطلب است که یک ارتباط علیتی بین دو متغیر وجود دارد به شکلی که میزان یکی (متغیر مستقل یا Independent) تا حدودی تعیین کننده دیگری (متغیر وابسته یا Dependent) است.
آزمون‌های آماری مورد استفاده جهت تجزیه و تحلیل اطلاعات با توجه به مقیاس اندازه گیری متغیرها به دو گروه پارامتری و ناپارامتری تقسیم می‌شوند. آزمون‌های پارامتری به تجزیه و تحلیل اطلاعات در سطح مقیاس فاصله‌ای و نسبی می‌پردازند که حداقل شاخص آماری آنها میانگین (Mean) و واریانس (Variance) است. در حالیکه آزمون‌های ناپارامتری به تجزیه و تحلیل اطلاعات در سطح مقیاس اسمی و رتبه‌ای می‌پردازند که شاخص آماری آنها میانه (Median) و نما (Mode) است.

انتخاب یک آزمون آماری :

انتخاب یک آزمون آماری وابسته به موارد زیر است:

۱) پرسش پژوهش
۲) نقشه یا طرح پژوهش
۳) نوع داده های جمع آوری شده

آزمون‌های پارامتری آمار استنباطی

آزمون t

آزمونt ، یک توزیع یا در حقیقت خانواده‌ای از توزیعهاست که با استفاده از آنها فرضیه‌هایی را درباره نمونه در شرایط ناشناخته بودن توزیع جامعه، آزمون می‌کنیم. اهمیت این آزمون (توزیع) در آن است که پژوهشگر را قادر می‌سازد با نمونه‌های کوچک (حداقل ۲ نفر) اطلاعاتی در باره جامعه بدست آورد. آزمون t شامل خانواده‌ای از توزیعها است و اینگونه فرض می‌کند، که هر نمونه‌ای دارای توزیع مخصوص به خود است، که شکل این توزیع از طریق محاسبه درجات آزادی (Degrees of Freedom) مشخص می‌شود. به عبارت دیگر توزیع t تابع درجات آزادی است و هر چه درجات آزادی (d.f) افزایش پیدا کند به توزیع طبیعی نزدیکتر می‌شود. هرچه درجات آزادی کاهش یابد، پراکندگی بیشتر می‌شود. خود درجات آزادی نیز تابعی از اندازه نمونه انتخابی هستند. هر چه تعداد نمونه بیشتر باشد بهتر است.
از آزمون t می‌توان برای تجزیه و تحلیل میانگین در پژوهش‌های تک متغیری یک گروهی و دو گروهی و چند متغیری دو گروهی استفاده کرد.

آزمون تحلیل واریانس

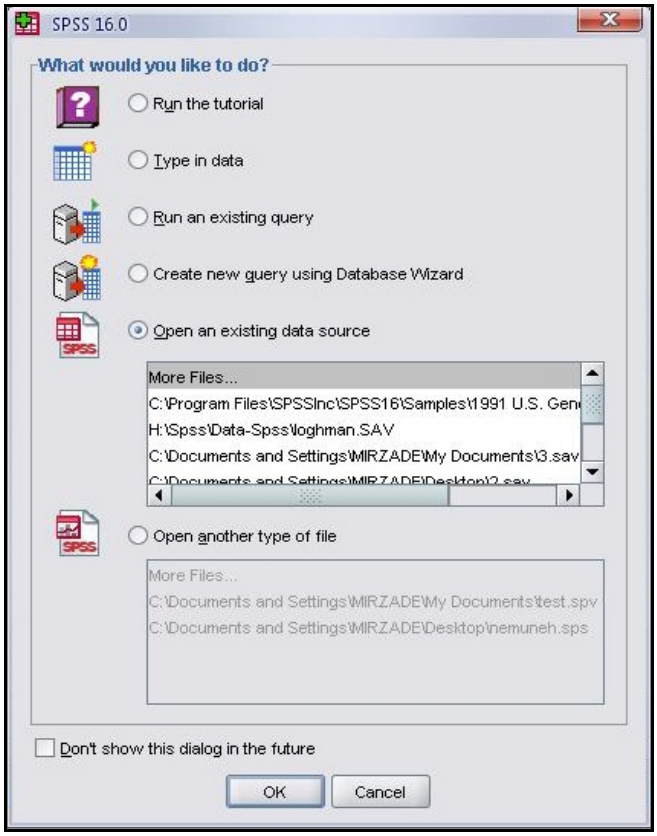
مواقعی که پژوهشگری بخواهد بیش از دو میانگین (بیش از دو نمونه) را مقایسه کند، باید از تحلیل واریانس استفاده کند. تحلیل واریانس یک روش فراگیرنده تر از آزمون t است و برخی پژوهشگران حتی وقتی مقایسه میانگین‌های دو نمونه مورد نظر است از این روش استفاه می‌کنند. طرح‌های متنوعی برای تحلیل واریانس وجود دارد و هر یک تحلیل آماری خاص خودش را طلب می‌کند. از جمله این طرح‌ها می‌توان به تحلیل یک عاملی واریانس) تحلیل یک طرفه) و تحلیل عاملی متقاطع واریانس، تحلیل واریانس چند متغیری، تحلیل کوواریانس یک متغیری و چند متغیری و .... اشاره کرد.

آزمون‌های ناپارامتری آمار استنباطی

در پژوهشهایی که در سطح مقیاسهای اسمی و رتبه‌ای اجرا می‌شوند، باید از آزمون‌های ناپارامتریک برای تجزیه و تحلیل اطلاعات استفاده شود. آزمون‌های زیادی برای این امر وجود دارد که براساس نوع تحلیل (نیکویی برازش، همسویی دو نمونه مستقل، همسویی دو نمونه وابسته، همسویی K نمونه مستقل و همسویی K نمونه وابسته) و مقیاس اندازه گیری می‌توان دست به انتخاب زد. از آزمون‌های مورد استفاده برای پژوهشها در سطح اسمی می‌توان به آزمون χ^2 ، آزمون تغییر مک نمار، آزمون دقیق فیشر و آزمون کاکرن اشاره کرد. از آزمونهای مورد استفاده برای پژوهشها در سطح رتبه‌ای می‌توان به دو آزمون کولموگروف- اسمیرونف، آزمون تقارن توزیع، آزمون علامت، آزمون میانه، آزمونu من ویتنی، آزمون تحلیل واریانس دو عاملی فریدمن و ... اشاره کرد.

شروع کار با SPSS

با اجرای نرم افزار SPSS محیط آن فعال شده و یک کادر محاوره ای ظاهر می‌شود. در این پنجره، نوع کاری که کاربر از SPSS میخواهد سئوال می‌شود.



اگر می خواهید با نحوه کارکردن با نرم افزار SPSS آشنا شوید، گزینه Run The Tutorial را انتخاب کنید .

اگر بخواهید داده‌های جدیدی را وارد SPSS کنید، گزینه Type In Data را انتخاب نمائید.

اگر می‌خواهید ازفایلهای بانک اطلاعاتی SPSS استفاده کنید، گزینه Run an exiting query را انتخاب نمائید.

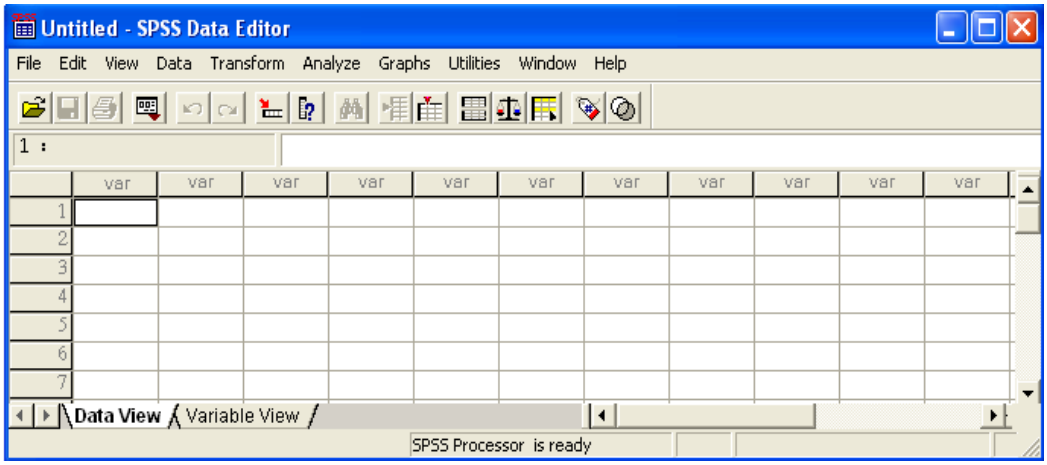
اگر بخواهید از فایلهای بانک اطلاعاتی تهیه شده در سایر نرم افزارها مانند SAS, dBase, EXEL, FOXPRO, Minitab و ...استفاده کنید، گزینه Create new query using database wizard را علامت دار کنید.

اگر بخواهید از مجموعه داده های ذخیره شده یا فایلهای داده ی SPSS استفاده کنید، گزینه Open an exiting data source را انتخاب کنید.

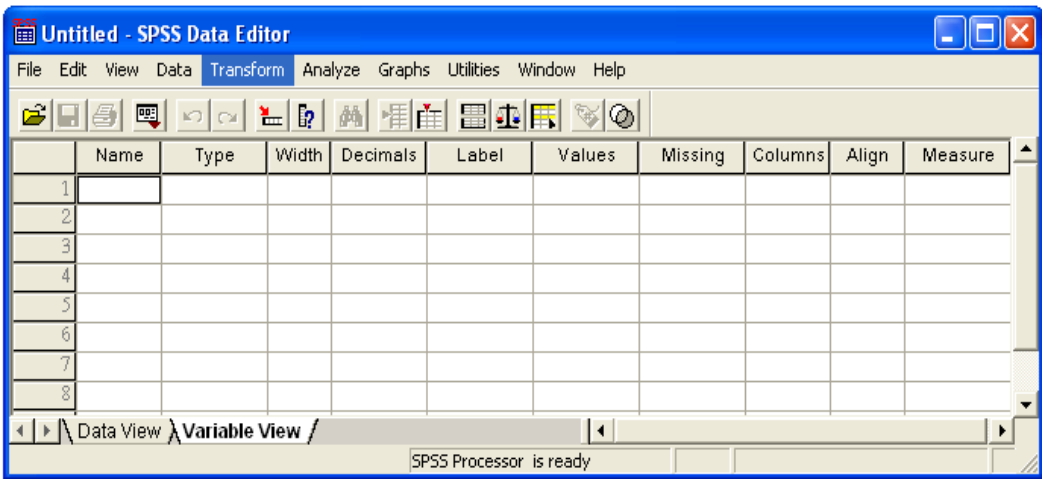
اگر بخواهیم سایر انواع فایلهای SPSS مانند خروجی ها را مشاهده و استفاده کنید، گزینه Open another type of file را انتخاب کنید.

اگر گزینه Don't show this dialog in the future را در کادر محاوره ای انتخاب کنید، در اجرای مجدد SPSS دیگر این کادر محاوره ای نمایان نمی شود و به صورت پیش فرض، یک صفحه داده خالی (Data sheet) به نام ***Untitled- SPSS Editor*** (ویرایشگر ***SPSS*** –بدون نام) نشان داده می شود که شامل دو پنجره مختلف به نامهای زیر می باشد:

۱ - Data view (پنجره ویرایش داده ها) :برای وارد کردن داده ها در زیر ستونهایی با نام پیش فرض Var و یا نامی که قبلاً معرفی شده است، استفاده می شود.



۲ - Variable view (نمای متغیر): برای تعریف کردن متغیرها و مشخصات هر متغیر استفاده می شود.



نوارهای SPSS

مانند همه پنجره ها در ویندوز همان طور که در شکل صفحه بعد می بینید، پنجره ویرایش گر داده ها دارای نوارهای زیر است.

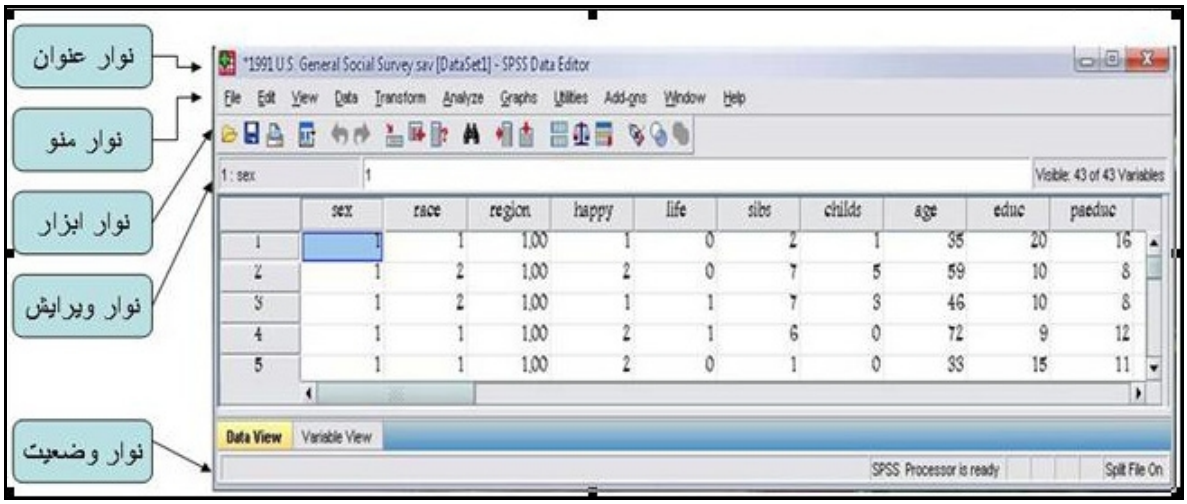
۱ نوار عنوان: در این نوار اسم فایل جاری و مشخصات آن نشان داده می شود.

۲ نوار منو: اصلی ترین نوار SPSS نوار منو است و تقریبا کلیه ی فعالیت های مربوط به باز و بسته کردن و ذخیره کردن فایلها، ویرایش، تجزیه و تحلیل داده ها و تغییرات در دستور اجرای نرم افزار، در گزینه های این نوار قرار دارند.

۳ نوار ابزار: برای دستیابی سریعتر به ابزارهایی که زیاد مورد استفاده قرار میگیرند، می باشد. در این نوار آیکون های این ابزارها را مشاهده میکنید. شما قادر خواهید بود نوار ابزار موجود را بدلخواه تغییر داده یا نوار ابزار جدیدی بسته به نیاز خودتان ایجاد کنید.

۴ نوار مخصوص Data Edit : این نوار شامل سه قسمت است. در سمت چپ میتوانید موقعیت هر سلول را مشاهده کنید. در بخش میانی میتوان مقادیر سلول فعال را مشاهده و ویرایش کرد. و در قسمت سمت راست تعداد متغیرهای فایل داده ها نمایش داده می شود.

۵ نوار وضعیت: در این نوار وضعیت موجود spss و فعالیتهای در حال اجرای آن نمایش داده میشود. مثلا اگر روی داده ها تغییراتی مانند Split یا Filter و مانند آن انجام داده اید می توانید در سمت راست این نوار وضعیت آنرا مشاهده کنید.

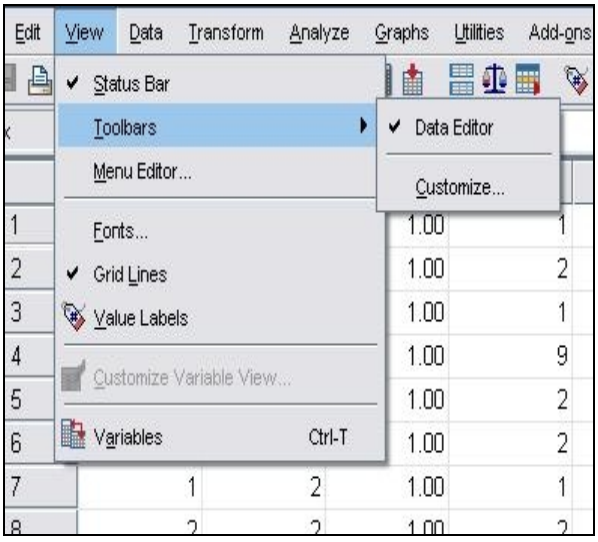


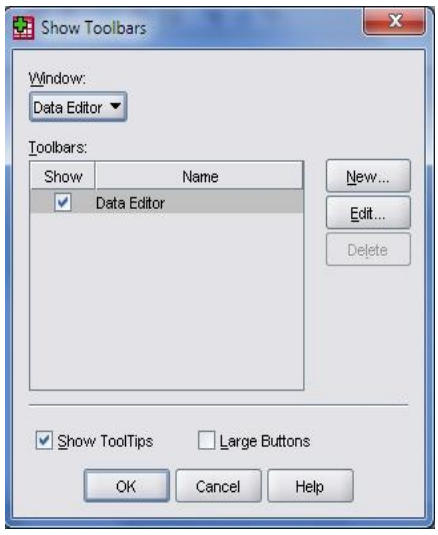
تنظیمات نوار ابزار:

می توانید بسته به نیاز خود ابزارهای نوار ابزار SPSS را کم یا زیاد کرده و یا اگر از شکل نمایش ابزارها خوشتان نمی‌آید آنها را تغییر دهید.

از نوار منو گزینه View و سپس گزینه Toolbar را انتخاب تا کادر محاوره ی (show toolbar) باز شود. ابتدا محیط کار خود را در قسمت Document Type مشخص کنید و هر ابزاری که می خواهید کم یا زیاد کنید.

برای تغییر نوار ابزار فعلی گزینه Edit… را انتخاب کنید و در کادر محاوره باز شده (شکل صفحه بعد) هر ابزاری که می خواهید اضافه یا کم کنید. برای ساختن نوار ابزار جدید از گزینه New… استفاده کنید.



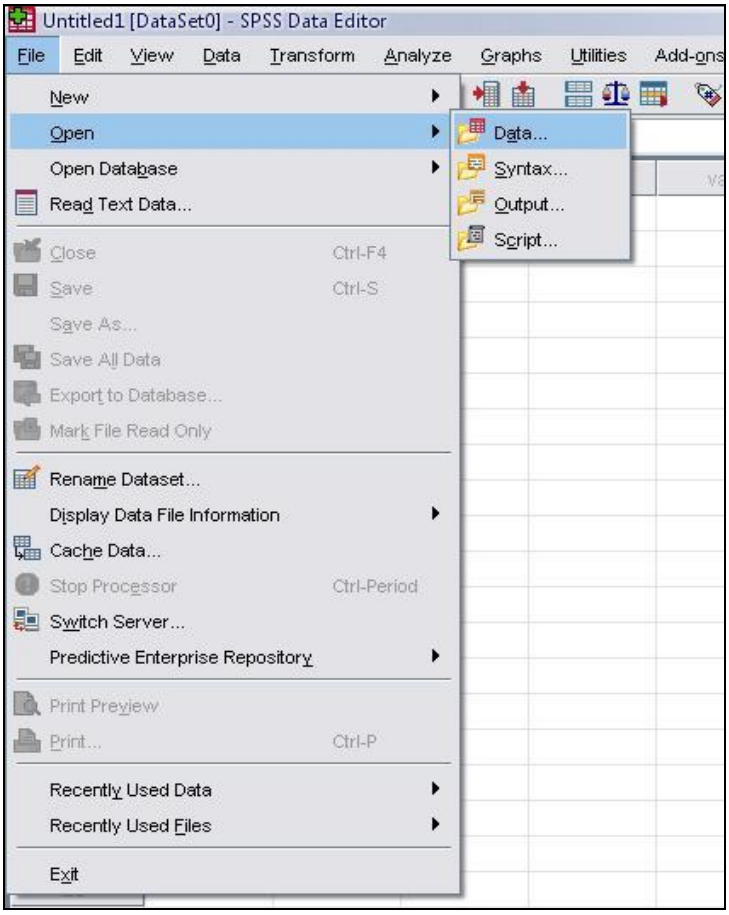


باز کردن یک فایل داده

برای استفاده از فایل داده های موجود در نوار منو، گزینه file را انتخاب کنید و بر روی گزینه open کلیک نمایید. با این عمل شما میتوانید فایلی را که می خواهید بر روی آن کار کنید باز کنید.

همچنین می‌توانید برای باز کردن یک فایل داده از کلید میانبر (Ctrl + O) و یا از آیکون Open File در نوار ابزار استاندارد SPSS استفاده کنید.

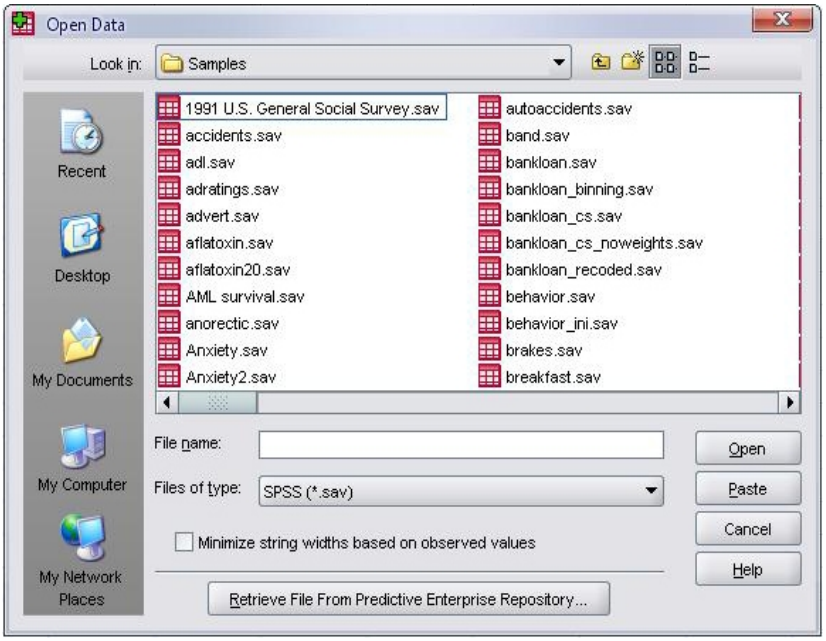
اگر می خواهید از فایل داده هایی که قبلا از آنها استفاده کرده اید مجددا استفاده کنید از نوار منو گزینه فایل recently use data را انتخاب و روی نام فایل مورد نظر خود کلیک کنید.



هنگام نصب SPSS تعدادی فایل داده به منظور استفاده کاربران در پوشه ای به نام sample در مسیر زیر قرار داده شده است.

C:\program file\spssinc\spss\sample

اگر نمی توانید فایل داده ها را پیدا کنید ، اطمینان حاصل کنید در دایرکتوری که Spss در آن نصب شده است فایل هایی با پسوند sav. موجود باشند.



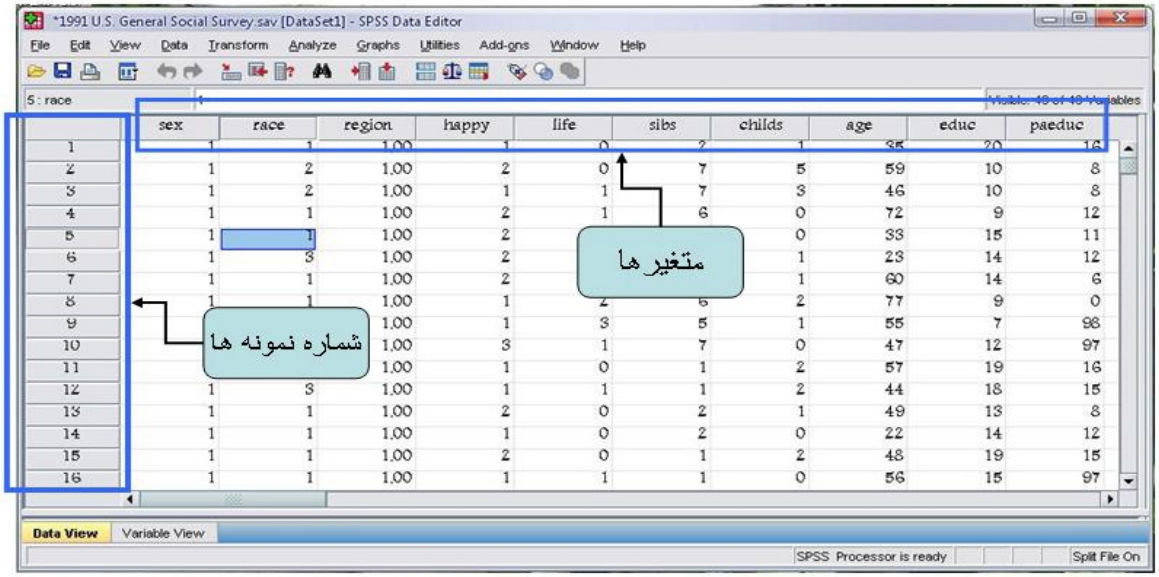
ویراشگر داده ها(Data Editor)

پنجره ای که در آن داده ها نمایش داده می شوند پنجره ویرایشگر داده ها (Data Editor) است که شامل دو پنجره فرعی به صورت زیر است:

۱- پنجره نمایش داده ها (Data View)

که محیط اصلی نمایش، وارد کردن و ویرایش داده ها است. این نمایشگر شامل تعدادی سطر (Case) که هر سطر اطلاعات مربوط به یک واحد نمونه را در بر دارد. بنابراین تعداد سطرها تعداد نمونه هایی است که از آنها اطلاعات جمع آوری شده است.

همچنین در این نمایشگر در بالای پنجره داده ها سطری است که نام متغیرها (Variables) آمده است. هر ستون یک متغیر را نمایش میدهد

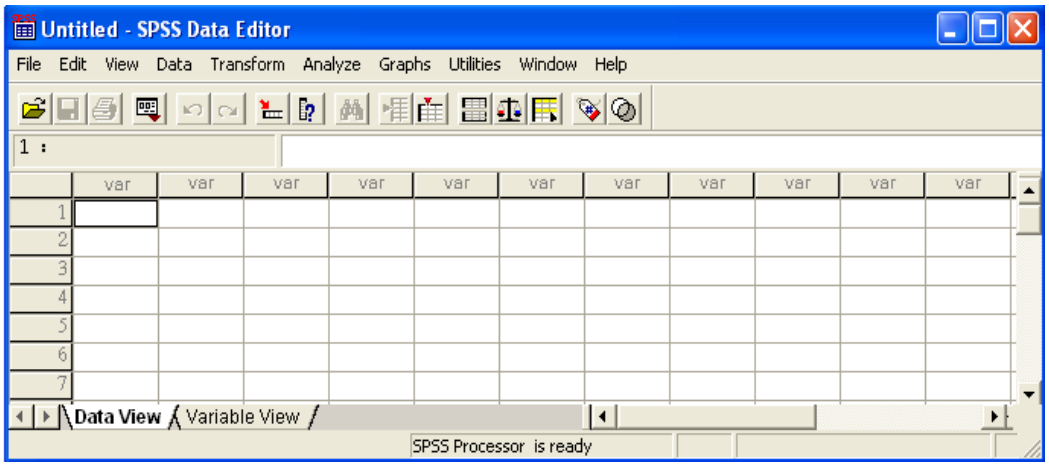


در نمایشگر داده ها هر مستطیل کوچک یک سلول است که داده ها در آن ثبت می شوند. برای وارد کـردن داده هـا کافی است روی هر سلول دوبار کلیک کنید و پس از وارد کردن داده با زدن Enter به سلول بعدی بروید.
راه دیگر، استفاده از نوار ویرایش گر سلول است که در بالای نام متغیر ها قـرار دارد و شـامل یـک بخـش بـرای وارد کردن داده ها و بخش دیگر موقعیت هر سلول فعال را نشان میدهد.
۲ پنجره نمایش متغیّر ها (Variable View) برای مشخص کردن نام ، نوع و سایر مشخصات هر یـک از متغیـر ها است. سطرها، تعداد نمونه ها را نشان می دهند. و در ستون های آن مشخصات متغیر ها نشان داده شده است.

وارد کردن متغیر ها

همانطور کـه قـبلاً اشـاره گردیـد وقتـی برنامـه ***Spss*** را بـاز مـی کنـیم پنجـره ای بـه نـام ***Untitled- SPSS Editor***(ویرایشگر ***SPSS*** – بدون نام) نشان داده می شود که شامل دو پنجره مختلف به نامهای زیر می باشد:

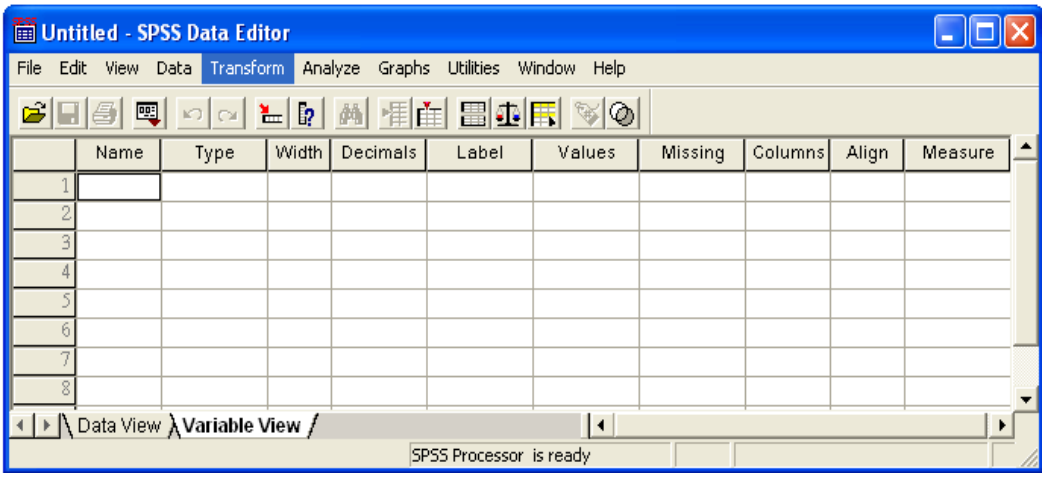
1- ***Data view*** (نمای داده) : برای وارد کردن داده ها و در زیر ستونهای معرفی شده استفاده می شود.



2- ***Variable view*** (نمای متغیر): برای تعریف کردن متغیرها استفاده می شود.

در حالت کلی، روش معمول این است که ابتدا با اسـتفاده از پنجـره ***Variable view*** متغیرهـا را تعریـف کـرده و سپس با استفاده از پنجره ***Data view***، اعداد را به SPSS وارد کنیم.

ابتدا از منوی اصلی گزینه فایل را انتخاب و پس از انتخاب گزینه New Data ، SPSS را بـرای ورود داده هـا و متغیر ها آماده کنید. سپس از برگه Variable View (مانند شکل زیر) برای ثبت مشخصـات متغیـر هـا اسـتفاده نمایید.



در ستون ***Name*** ابتدا نامی برای متغیر انتخاب می کنیم (در اینجا داده های مربوط بـه معـدل دانشـجویان)،بعـد از وارد کردن نام، سایر ستونها با پیش فرضهایی که نرم افزار طراحی کرده است به صورت زیر نمـایش داده مـی شـوند که ما می توانیم به دلخواه و با توجه به نوع داده، تغییراتی در ستونها ایجاد کنیم.

- روی سلول زیر Name نام متغیر دو بار کلیک کنید و سپس نام مورد نظرتان را وارد کنید.

قواعد نام گذاری متغیر ها

انتخاب نام متغیر ها در spss تابع قوانین زیر است:

۱ مجاز هستید تا ۶۴ کاراکتر برای نام متغیر اختصاص دهید.

۲ با حرف شروع شود، اما نباید از فضای خالی و کاراکترهای نظیـر ؟، " ، ! و کلمـات کلیـدی SPDD(BY.ALL AND,NOT,EQ,) استفاده شود.

نام متغیر می تواند شامل حروف کوچک یا بزرگ، عدد یا یکی از کاراکتر های @ و# و . و _ و $ باشد.

(توجه: باید از کاراکتر های # و . و _ و $ فقط در بین نام متغیر ها استفاده کنید.)

۳ از گذاشتن فاصله در نام متغیر خود داری کنید.

۴ از گذاشتن کاراکتر های # و $ در ابتدای نام یک متغیر اجتناب کنید.

۵ نام متغیر میتواند با @ شروع شود .

۶ نام متغیر نباید با کاراکتر های . یا _تمام شود. یعنی در انتهای نام نقطه گذاشته نشود.

۷ نام متغیر نمیتواند تکراری باشد.

۸ نام متغیر نباید یکی از کلمات کلیدی مانند

NOT, OR, TO, WITHALL, AND, BY, EQ, GE, GT, LE, LT, NE ، کـه spss بـه عنـوان عبارت محاسباتی از آنها استفاده می کند، باشد .

انواع متغیر:

تعریف متغیر (**Variable**) : ویژگی یا خاصیت یک فرد، شئ و یا موقعیت است که بـرای هـر فـرد مقـادیر مختلفـی بخود می گیرد. قد، وزن، گروه خونی و جنس مثال هایی از متغیرها هستند. متغبرها ماهیتاً دو گونه می باشند.

۱ متغیرهای ***کمی*** که قابل اندازه گیری و درجه بندی می باشند. مانند قد، وزن یا سن.

۲ متغیرهای ***کیفی*** کـه فقط قابلیت تفکیک در رده های مختلف هسـتندو قابـل بیـان بـا اسـتفاده از واحـد خاصـی نیستند. مانند جنس، گروه خونی یا ملیت.

داده (**data**) : داده عبارتست از نمایش ذخیره شده ویژگیها و مشخصات قابل مشاهده واحدهای نمونه ای اسـتخراج شده از جامعه آماری.

انواع داده :

۱) داده های ***کمی*** (فاصله ای) Interval اعدادی هستند که بیانگر کمیـت بـه صـورت واحـدهای عـددی و بـر اساس یک مقیاس مستقل است. قد و وزن مثالهای بارز داده های کمی هستند.

۲) داده های ***رتبه ای*** Ordinal مشتمل بر رتبه ها، تعلق داشتن به گروههـای رتبـه بنـدی شـده یـا اطلاعـات ترتیبی است. به عنوان مثال اگر دو داور به یک مجموعه ۱۰ تایی از نقاشی رتبه یک (برای بهترین) تـا رتبـه ۱۰ (برای بدترین) بدهند. مجموعه داده ها مشتمل بر ۱۰ جفت رتبه خواهد بـود. کـه هـر جفـت بـرای یـک نقاشی است.

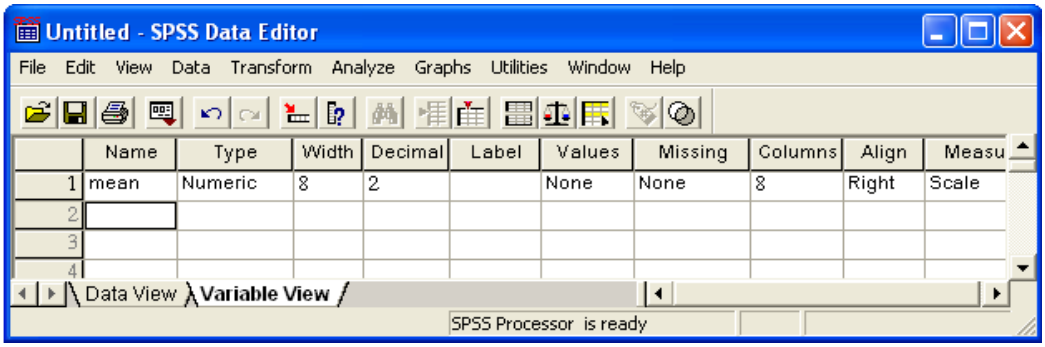
۳) داده های ***اسمی*** (nomial) که مربوط به متغیر یا خواص کیفی مانند جنس یا گروه خونی است و بیانگر عضویت در یک گروه خاص می باشد.

مثال: معدل دیپلم(کمی) و نوع دیپلم(کیفی) ۱۵ نفر از دانشجویان سال اول دانشگاه در چند رشته مختلف دانشگاهی به صورت زیر می باشد. به منظور استفاده از این اطلاعات در تجزیه و تحلیلها، اطلاعات (داده ها) را به صورت زیر به نرم افزار وارد می کنیم.

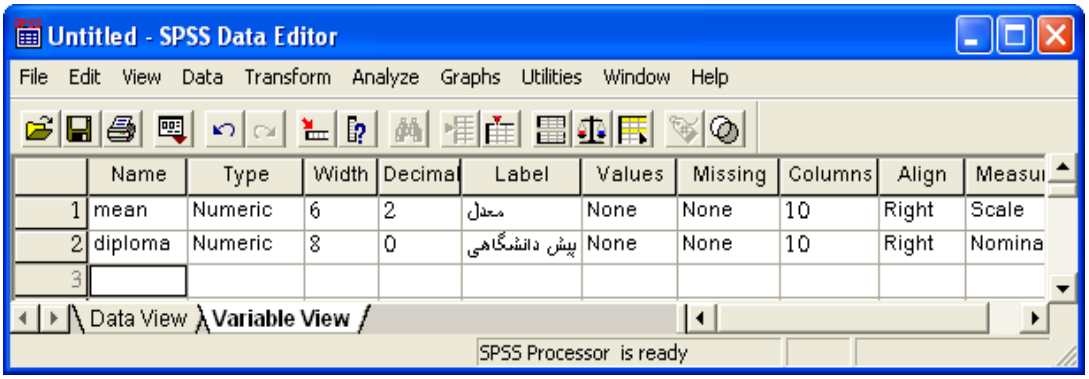
کد ۱ = دیپلم ریاضی کد ۲ = دیپلم تجربی کد ۳ = دیپلم انسانی

دیپلم	1	2	2	2	3	1	1	3	3	3	2	2	1	1	2
معدل	17	16.5	18.3	17.89	15.6	17.3	14	16.7	17.8	16	18.8	17.9	17.6	18	19

برای این کار بعد از وارد شدن به نرم افزار SPSS به صورت زیر عمل می کنیم:



به همین ترتیب متغیر نوع دیپلم را در پنجره ***Variable view*** تعریف کرده و متناسب با نوع داده، ستونهای مختلف را تنظیم می کنیم.



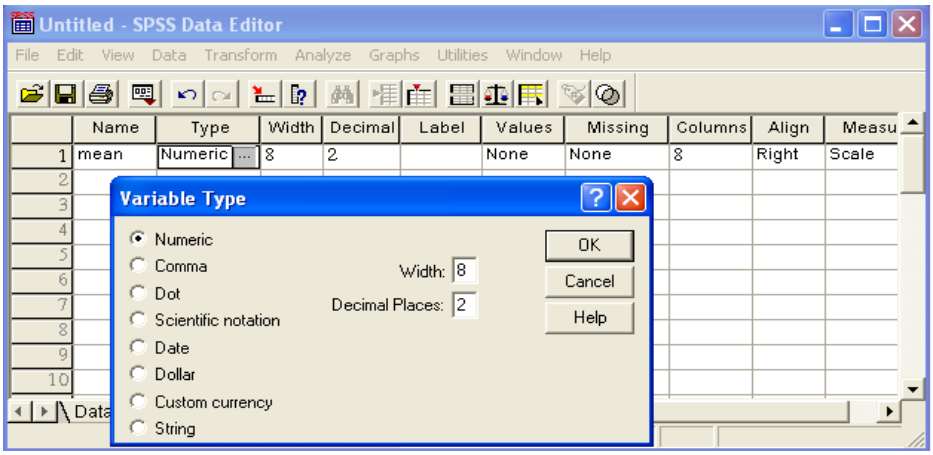
بدین صورت که، درستون Type نوع متغیر را تعیین می کنیم. با کلیک بر روی Type (نوع متغیر) سه نقطه خاکستری رنگ روی سلول آن نمایش داده میشود که اگر روی آن کلیک کنید پنجره Variable Type باز شده و به صورت پیش فرض نوع متغیر از نوع عددی (Numeric)انتخاب شده است. در پنجره Variable Type از بین گزینه ها نوع داده مناسب را برای متغیر مورد نظر تعیین می کنیم. مثلاً برای معدل دانشجویان که از نوع عددی می باشد همان گزینه Numeric (داده عددی) را بدون تغییر باقی می گذاریم.

اگر متغیر از نوع عددی نیست گزینه String را علامت دار کنید. و اگر چیزی غیر از اینها مثل تاریخ، نماد علمی یا پول و ...است، میتوانید یکی را انتخاب کنید.

گزینه Comma (ویرگول) در آن هر عدد سه رقم سه رقم با یک ویرگول جدا می شوند.

گزینه Dot(نقطه):ارقام سه رقم سه رقم با نقطه از هم جدا می شوند.

گزینه Scientific notation(نماد علمی): اگر این گزینه انتخاب شود اعداد با نماد علمی نشان داده می شوند.مانند عدد ۴۸۰۰۰۰ که بصورت ۵ E+۴۰۸ نمایش داده می شود.

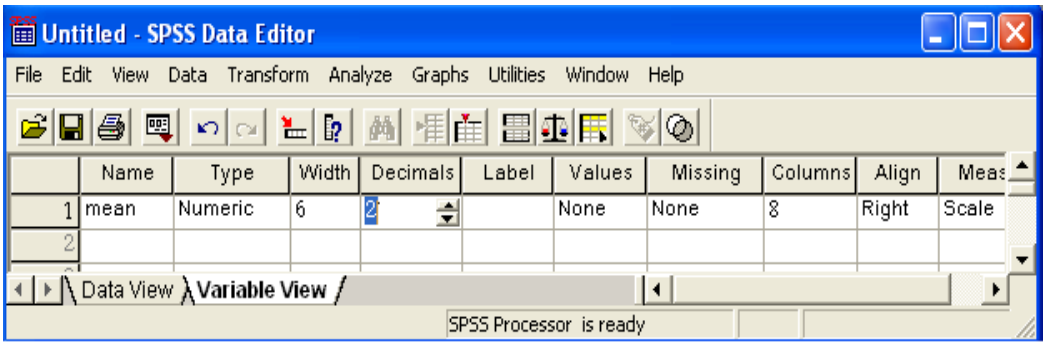


اگر داده ها اعشار دارند در قسمت Width تعداد ارقام صحیح و در قسمت Decimal تعداد ارقام اعشار را مشخص کنید.

پهنای متغیرهای عددی با کلیک روی ستون Width، دو پیکان کوچک بالا و پایین نشان داده می شود که می توانیم با بالا و پایین کردن، پهنای مورد نظر را تغییر دهیم.

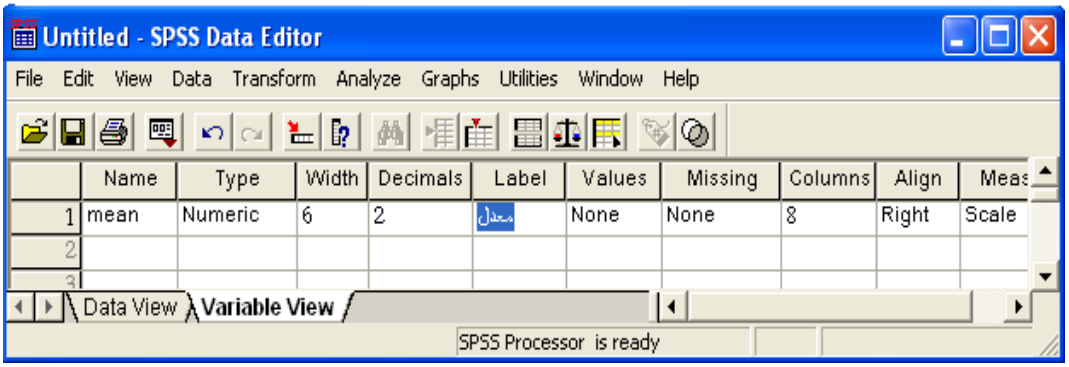
از ستون Decimals برای تعیین تعداد ارقام اعشار داده های مورد نظر استفاده می شود. مانند روش قبل روی ستون کلیک کرده و تعداد ارقام اعشار را تعیین می کنیم.

Widthو Decimals - تعداد اعداد صحیح و اعشار را میتوانید با کلیک روی آنها و افزایش یا کاهش آن به دلخواه تغییر دهید .برای این داده ها با توجه به اینکه معدل یک فرد می تواند به طور مثال ۱۸.۲۵ باشد به همین دلیل ستون مورد نظر را به همان صورت پیش فرض نگه می داریم. به این معنی که در ستون معدل در پنجره (Data View) اعداد مربوط به معدل دانشجویان دو رقم اعشار نشان داده می شوند.

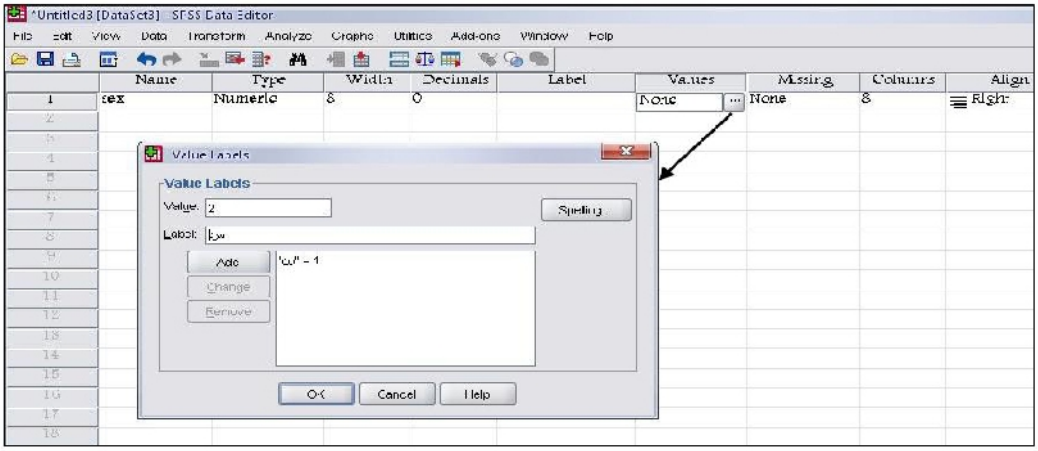


- در ستون Lable می توان برای متغیر مورد نظر یک برچسب انتخاب کرد. با انتخاب بر چسب، میتوانید برای شناخت بهتر متغیر از تعاریف مناسب استفاده کنید و تا ۱۲۰ کاراکتر وارد کنید.

برای این کار در ستون Label عنوان مورد نظر را تایپ می کنیم.

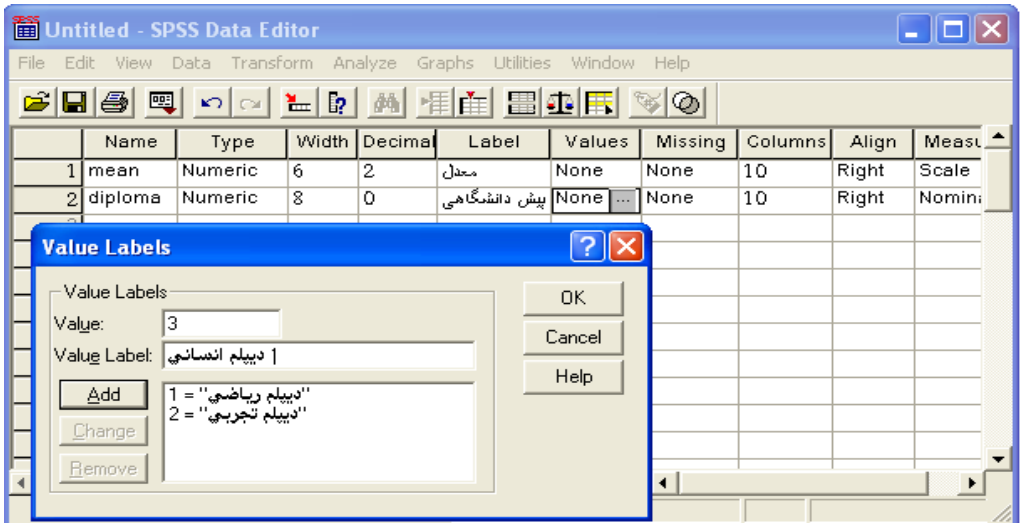


- ستون مربوط به ***Values*** (Values labels) برای متغیرهای گروه بندی مورد استفاده قرار می گیرد. در مورد متغیر های کیفی و یا گروه‌بندی شده لازم است که گروهها و حالتهای مختلف متغیر را با مقادیر و بر چسب هایی مشخص کنید. پس از کلیک بر روی این مشخصه با سه نقطه خاکستری ظاهر شده که با کلیک روی آن پنجره value labels ظاهر شده و میتوان به هر گروه و یا هر حالت از متغیر کیفی یک عدد نسبت داد. مانند روش زیر ابتدا کدهای مورد استفاده برای سطوح مختلف متغیر کیفی مورد نظر را وارد کرده، سپس در سطر مربوط به ***Value label*** عنوان مورد نظر را تایپ کرده و سپس روی گزینه ***Add*** کلیک می کنیم. بعد از وارد کردن تمامی کدها و عناوین مربوطه روی گزینه ***Ok*** کلیک می کنیم.



با توجه به اینکه نوع دیپلم داده اسمی می باشد، در این قسمت می توان در ستون ***Values*** هر کدام از کدهای نوع دیپلم را با یک برچسب نشان داد.

این کار باعث می شود که در خروجیهای ما به طور مثال به جای نمایش کد ۱ معادل آن یعنی دیپلم ریاضی مشاهده شود.

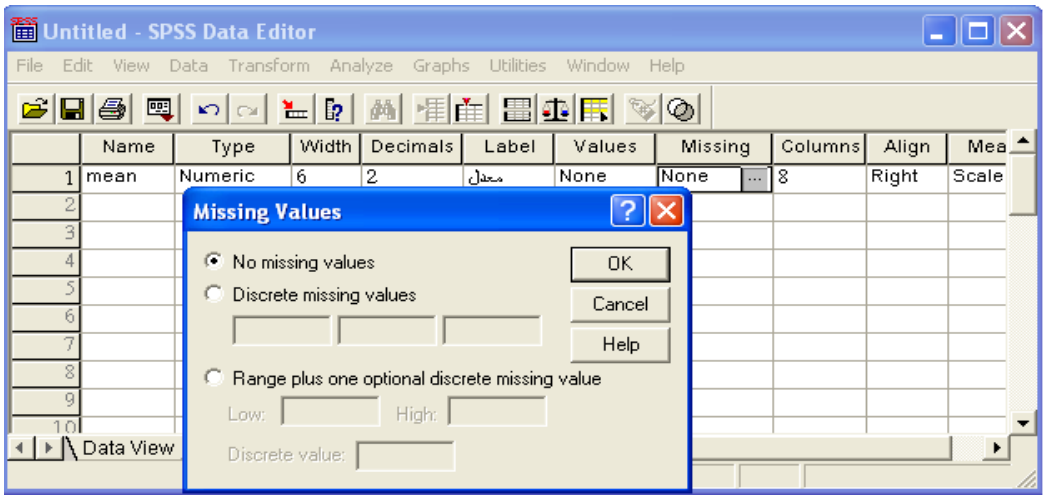


برای معدل دیپلم دانشجویان، چون متغیر گروه بندی نمی باشد، اطلاعاتی در این قسمت اضافه نمی کنیم.

- ستون بعدی (Missing) مربوط به داده های گمشده می باشد. در این قسمت با کلیک بر روی ستون مورد نظر پنجره Missing Values باز می شود.

Missing مقادیر گم شده ای هستند که به هر دلیلی یا ثبت نشده اند، یا اشتباه ثبت شده اند و یا پاسخ دهنده از پاسخ به آن امتناع کرده است. در چنین حالتی کاربر می‌تواند هنگام ثبت داده ها برای هریک از این حالت‌ها عدد یا کاراکتری را در نظر گیرد.

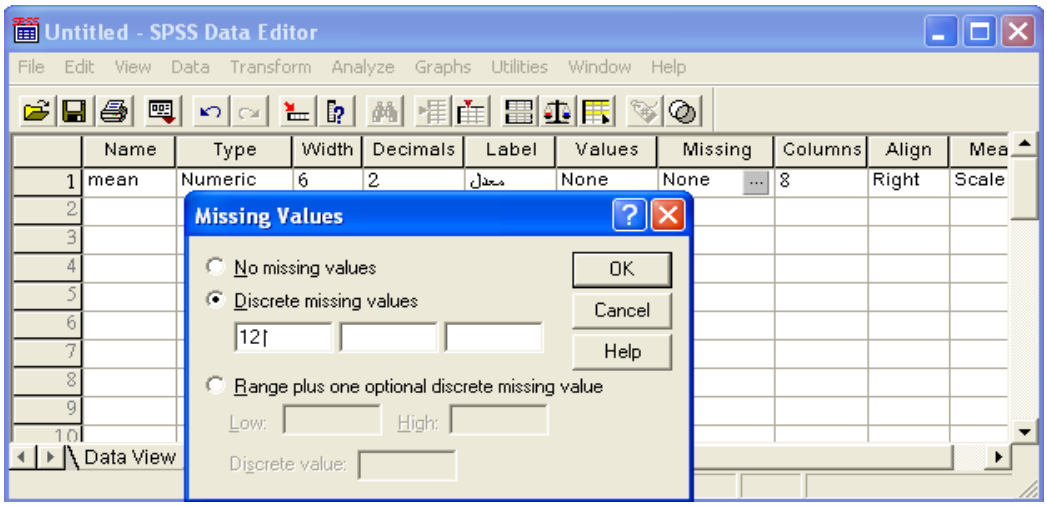
برای این کار پس از کلیک بر روی سلول زیر این مشخصه، سه نقطه خاکستری ظاهر شده که با کلیک روی آن پنجره مربوط به Missing value ظاهر شده و می‌توان برای داده های گم شده مقادیری غیر از مقادیر متغیر در نظر گرفت.



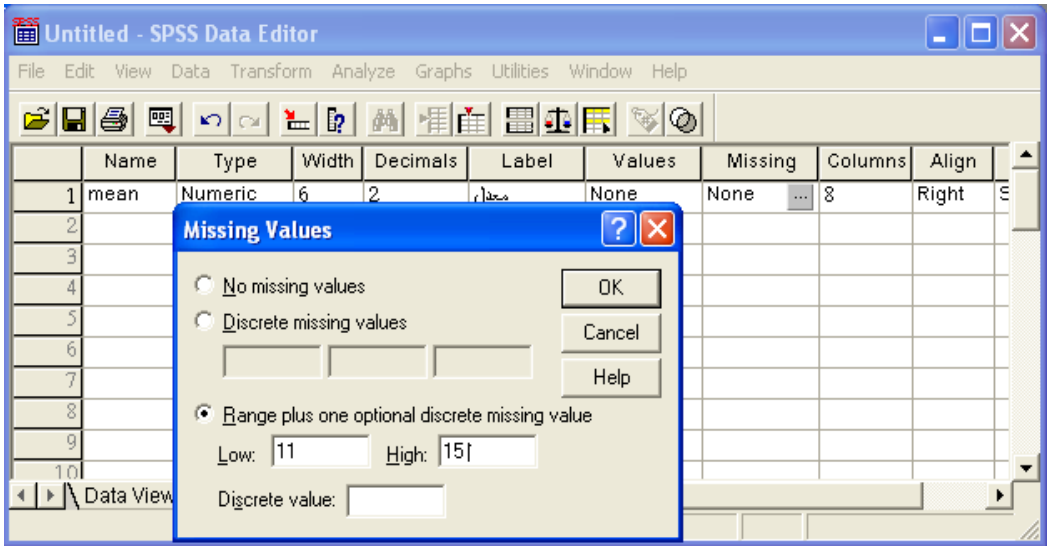
زمانی که در بین اطلاعات جمع آوری شده داده گمشده ای وجود نداشته باشد پیش فرض ***No missing values*** به همان حالت خود باقی می ماند. اما اگر داده گمشده وجود داشته باشد، برای مشخص کردن آن باید گزینه ***Discrete missing values*** را فعال کرد و شماره سطر مربوط به داده مورد نظر را در مستطیلهای زیر وارد کرد.

به طور مثال اگر معدل یکی از ۱۵ دانشجو در دسترس نباشد، به صورت زیر عمل می کنیم:

اگر عدد، مربوط به خانه شماره ۱۲ که بیان کننده معدل دانشجوی دوازدهم است، باشد و به بیان دیگر داده گمشده باشد باید شماره ۱۲ را به صورت زیر در مربع مورد نظر وارد کرد. به همین ترتیب اگر داده گمشده دیگری داشتیم، شماره های آنها را در مستطیلهای بعدی وارد می کنیم. (برای حداکثر۳ داده گمشده در مستطیلهای بالایی)

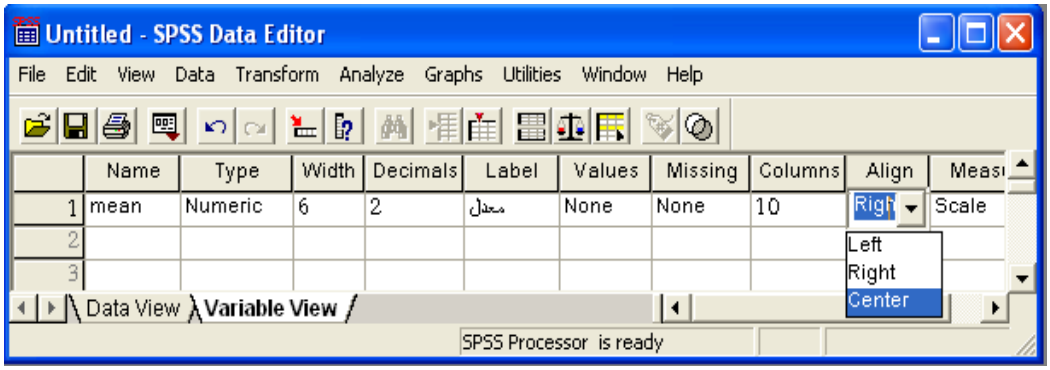


اگر تعداد بیشتری داده گمشده وجود داشته باشد، داده ها را به ترتیب کوچک یا بزرگی تنظیم می کنیم، سپس با فعال کردن قسمت ***Range plus one optional discrete missing value*** در مستطیلهای پایینی شماره های داده گمشده را مشخص می کنیم. (از شماره .... تا شماره....) که به طور مثال در زیر نشان داده شده است.



و اگر داده های گمشده به صورتی بودند که یک سری از آنها پشت سر هم و یکی از آنها جدا بود، شماره داده گمشده جدا را در قسمت ***Discrete value*** اضافه می کنیم.

- ستون بعدی ***Columns*** مربوط به تغییر دادن پهنای ستون در پنجره ***Data view*** می باشد؛ استفاده از این گزینه روش دیگری برای تغییر تعداد کاراکتر های یک متغیر است.
- تراز کردن دادهها در ستون ***Align*** قابل انجام شدن می باشد.
- Align برای راست نویسی، چپ نویسی و وسط نویسی داده ها در سلولها از این مشخصه استفاده کنید.



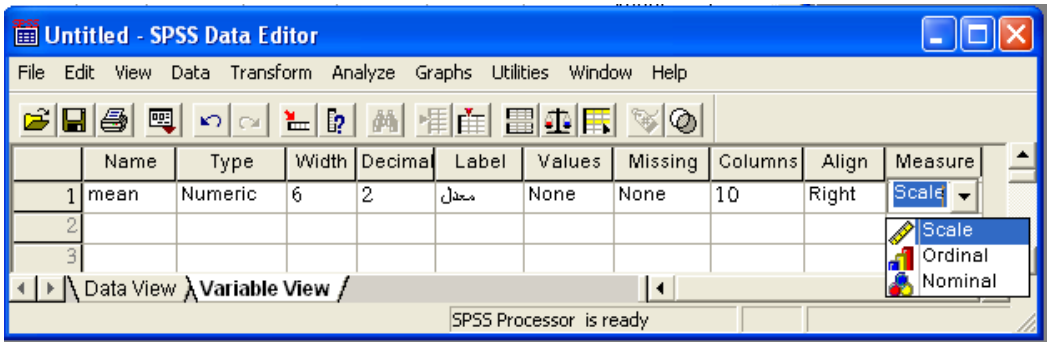
داده ها به طور پیش فرض در پنجره ***Data view*** راست چین هستند ولی زمانی پیش می آید که می خواهیم داده ها چپ چین و یا وسط چین باشند. برای این کار ابتدا مکان نمای ماوس را روی ستون ***Align*** قرار داده و گزینه مورد نظر را انتخاب می کنیم.

- آخرین ستون مورد بررسی در پنجره ***Variable view*** مربوط به ستون مقیاس اندازه گیری متغیرها (***Measure***) می باشد. که شامل سه نوع زیر است:

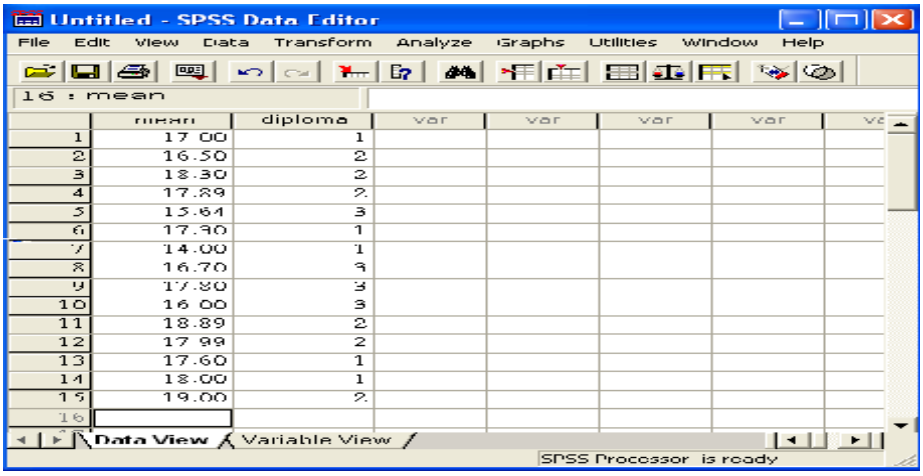
از مهمترین مشخصه های یک متغیر تعیین مقیاس اندازه گیری آن است که در این قسمت تعیین می شود. این مشخصه دارای سه گزینه است که مقیاس اندازه گیری هر متغیر را معلوم می کند.

۱ Nominal را برای مقیاس اسمی ۲ Ordinal را برای مقیاس ترتیبی ۳ Scale را برای مقیاس عددی بکار ببرید.

مثلاً داده های مربوط به معدل دانشجویان از نوع ***Scale*** می باشند.



پس از تعریف متغیرها ، در پنجره ***Data view*** اعداد را به صورت زیر وارد می نماییم.



به ۳ روش میتوانید داده ها را به SPSS منتقل کنید.

۱ تایپ کردن داده ها در نوار ویرایش داده ها.

۲ با کلیک کردن برروی هر سلول و وارد کردن داده ها در داخل سلولها.

۳ کپی کردن از مکان دیگر و چسباندن در فایل داده های موجود.

انواع فایلها در SPSS

۱ فایل داده‌ها (data file):

این فایلها ها که با پسوند sav نمایش داده می شوند، حاوی اطلاعاتی است که کاربر در پنجره Data View (محیط ویرایش داده ها) آنها را به نرم افزار وارد می کند. با نحوه ایجاد اینگونه فایلها در بخش قبلی آشنا شده اید.

۲ فایل‌های برنامه نویسی (syntax file):

این فایل ها با پسوند sps. مشخص میشوند و محیطی برای اجرای دستورات spss هستند تا اگر کاربر با استفاده از امکانات موجود در پنجره ها نمیتواند نیاز خود را مرتفع کند با نوشتن یک برنامه با زبان spss یک فایل syntax ایجاد کرده و پس از اجرای آن، تحلیل مورد نیاز خود را انجام دهد.

۳ فایل‌های script:

با پسوند sbs. مشخص می شوند. کاربر میتواند با اجرای این گونه فایل های از پیش ساخته شده، محیط و پنجره های نرم افزار را به دلخواه تغییر دهد.

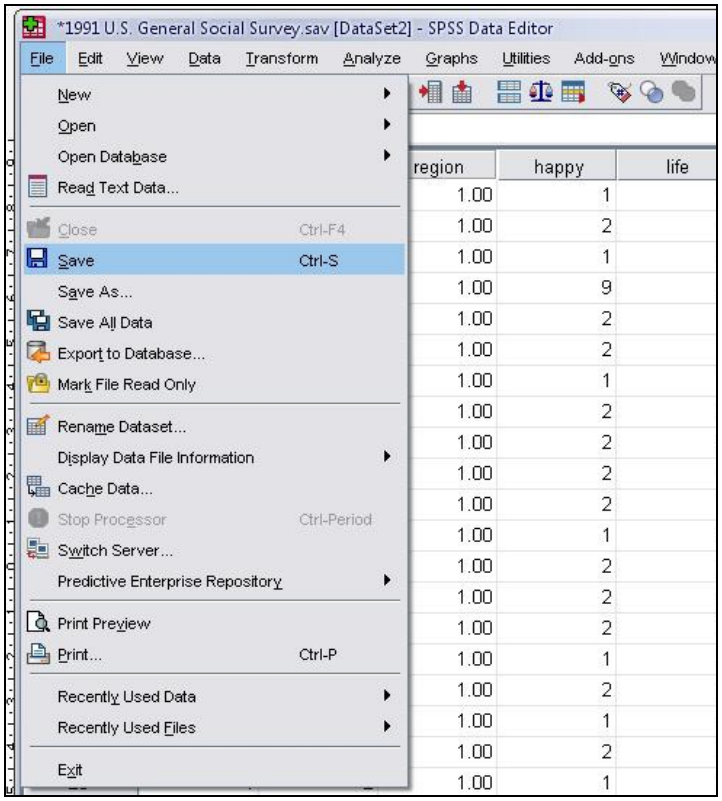
۴ فایل‌های خروجی (Output file):

از ابتدا تا پایان کار با نرم افزار، کلیه عملیاتی که در spss انجام میدهید در یک فایل خروجی نگهداری میشود. هنگام خروج از نرم افزار در مورد نگهداری این فایلها از شما پرسیده میشود، در صورت تمایل، میتوانید آنها را با نام دلخواه ذخیره کنید. این گونه فایل ها با پسوند spo مشخص میشوند و قابل ویرایش هستند. چون همه نتایج کار خود را میتوانید در این فایلها ذخیره کنید، با نحوه کار با فایلهای خروجی بیشتر آشنا خواهید شد.

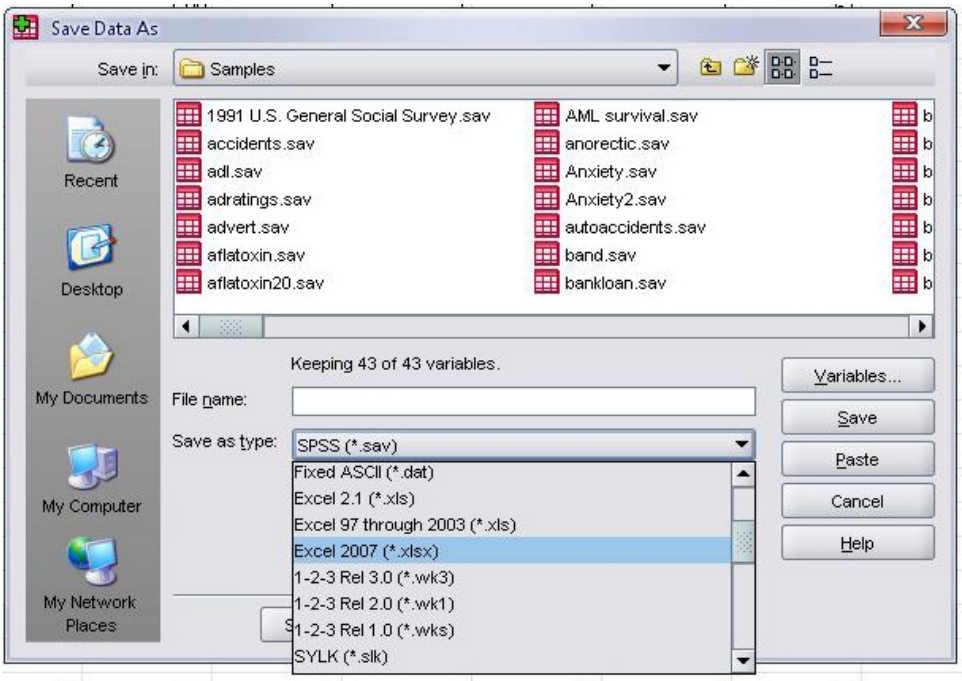
ذخیره کردن فایلها

چون زمان زیادی صرف وارد کردن داده ها می شود، اگر حجم بزرگی از داده ها را وارد می کنید هر چند دقیقه یک بار ذخیره سازی توصیه میشود. برای ذخیره کردن فایلها در SPSS برای اولین بار همانند تمام برنامه‌های تحت ویندوز از منوی اصلی، فایل و گزینه Save As را انتخاب و با وارد کردن یک نام برای فایل آنرا ذخیره کنید.

برای ذخیره سازی مجدد می‌توانید از کلید میانبر (Ctrl +S) و یا از آیکون Save File در نوار ابزار SPSS برای استفاده کنید.



اگر در هنگام ذخیره کردن فایل داده ها بخواهید آن را با فرمت دیگری برای استفاده فایل داده ها در سایر برنامه ها مانند Excel می توانید در منوی اصلی گزینه File دستور Save as را انتخاب کنید و در پنجره باز شده مانند شکل نامی را برای فایل خود انتخاب و سپس در پایین همان پنجره فرمت مورد نظرتان را انتخاب کنید و کلید ok را فشار دهید.

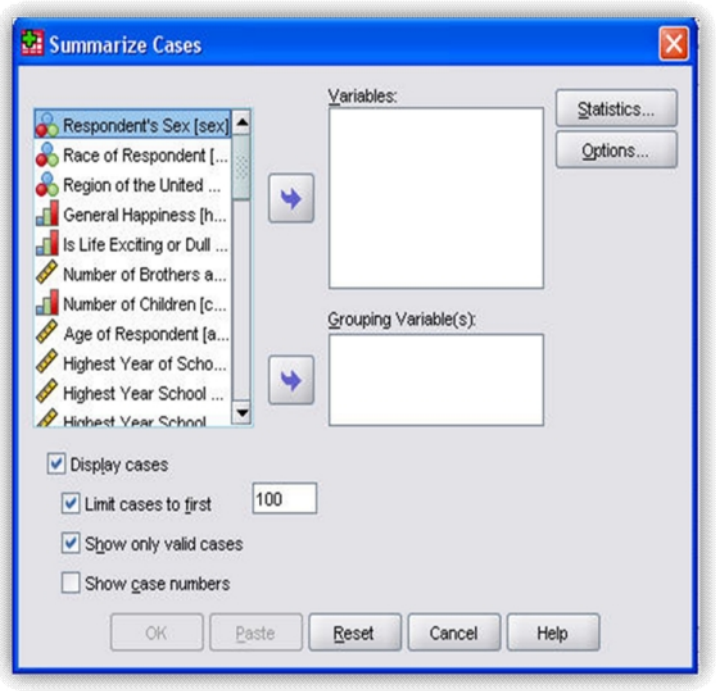


تهیه گزارش از فایل داده ها

گاهی ممکن است به منظور خاصی بخواهید فایل داده ها را به صورت یک فایل خروجی بصورت Text در اختیار داشته باشید. به عنوان مثال اگر بخواهید از فایل داده ها، چند متغیر را انتخاب و اطلاعات مربوط به نمونه ها را در خروجی مشاهده کنید و به صورت یک گزارش آنها را چاپ کنید، دستورزیر را از منوی Analyze اجرا کنید:

Analyze/Report/Case Summaries…

تا کادر محاوره Summarize Cases مانند شکل باز شود.



در این کادر محاوره متغیرهایی را که می‌خواهید فهرست کنید. از کادر سمت چپ به کادر Variables منتقل کنید.

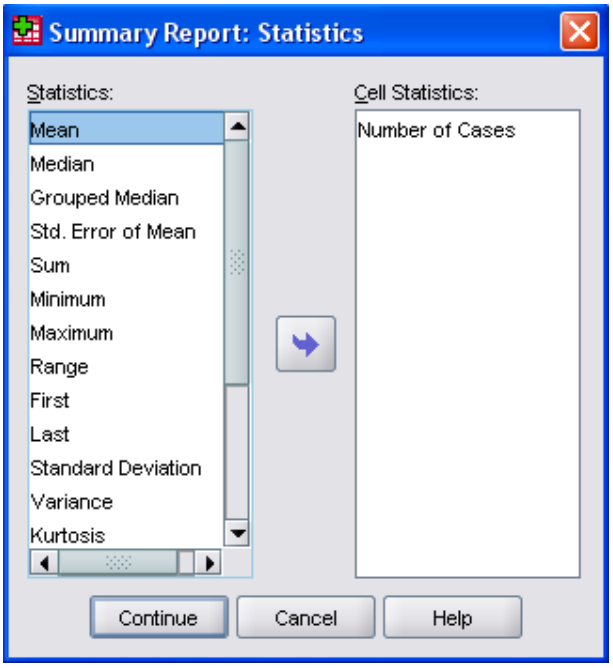
اگر می خواهید اطلاعات مربوط به همه نمونه ها و یا بخشی از آنها را در خروجی داشته باشید گزینه Display Cases را علامت دار کنید تا همزمان با آن گزینه های بعدی نیز فعال شوند.

اگر می خواهید تعداد محدودی از داده ها را لیست کنید در کادر مربوطه به گزینه Limited Cases To First تعداد مورد نظر را وارد کنید. مثلا اگر عدد ۱۰ را وارد کنید. ۱۰نمونه اول از شماره ۱ الی ۱۰ را در خروجی مشاهده خواهید کرد.

اگر گزینه Show Only Valid Cases را تیک بزنید داده های گم شده ، در خروجی ظاهر نمی شوند.

انتخاب گزینه Show Case numbers باعث می شود که شماره نمونه ها نیز در فهرستی که تهیه می کنید، مشاهده شود.

اگر قصد دارید برای متغیرهایی که انتخاب کرده‌اید، بعضی از آماره‌های توصیفی را محاسبه کنید، از گزینه Statistics استفاده کنید تا کادر محاوره آن باز شود.



در کادر مکالمه آن می‌توانید آماره های توصیفی دلخواه را به کادر Cell Statistics: (سمت راست) منتقل کنید.

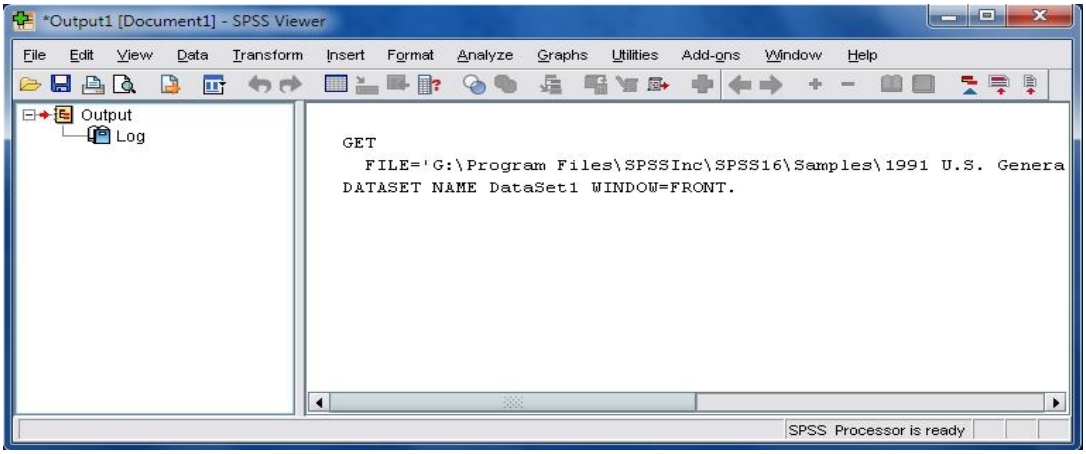
دقت کنید آماره‌های انتخابی برای متغیرهایی که انتخاب کرده‌اید، با معنی باشند.

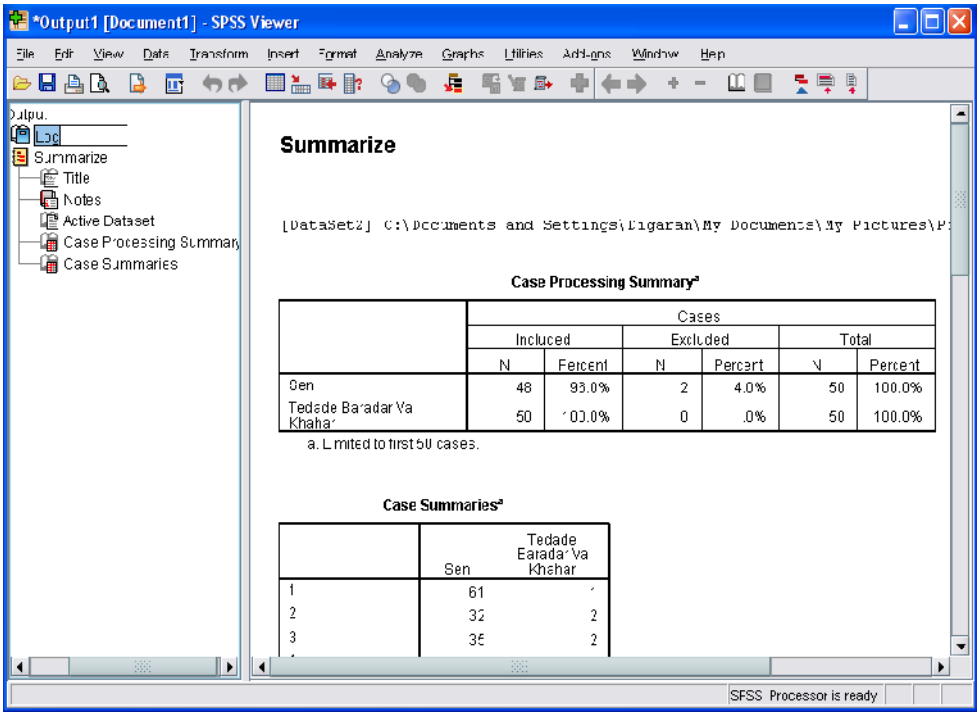
برای ادامه Continue و Ok را کلیک کنید.

اینک می‌توانید نتیجه را در خروجی مشاهده کنید.

پنجره خروجی ها

هنگامی که شما برنامه SPSS را باز می کنید، به طور همزمان دو پنجره باز می شود. یک پنجره Data Editor که پنجره اصلی SPSS است و پنجره دیگر پنجره خروجی ها Viewer است. این پنجره مانند یک رابط بین نرم افزار و کاربر است. شما هر عملیات و دستوری را که در نرم افزار انجام می دهید حاصل عملیاتی که انجام می شود در خروجی به شما گزارش می شود. از ابتدا که با نرم افزار شروع به کار می‌کنید تا زمانی که قصد خروج دارید ، نتایج همه فعالت‌های شما در خروجی ثبت می‌شود. به اشکال زیر توجه کنید.





پنجره خروجی Output مکانی است که نتیجه کارتان را در آن مشاهده می‌کنید و شامل دو بخش است. بخش Outline Pane که نمای کلی اجزای خروجی یا سرفصل تمام نتایج موجود از Viewer را به نمایش می‌گذارد و بخش دوم در سمت راست و به نام Display Pane است که در آن اشیاء خروجی به نمایش گذاشته شده است. هر بخش از خروجی مانند یک جدول یا نمودار، با یک آیکون در سمت چپ Outline Pane در ارتباط است و هر آیکون نماینده قسمتی از خروجی است.

در Outline Pane کتابچه هایی را مشاهده می کنید. با کلیک بر روی هریک از کتابچه‌ها در سمت چپ شیئ مربوط به آن در Display Pane قابل رؤیت است.

یک کتاب بسته نشان‌دهنده آن است که آیتم مربوط به آن در خروجی فعلاً قابل رؤیت نمی باشد.

این پنجره در واقع نتایج تجزیه و تحلیل داده ها در SPSS می باشد. برای ذخیره و بازیابی این فایل همانند سایر نرم افزارهای کاربردی از منوی File دستور ذخیره و یا بازیابی استفاده می گردد.

آشنایی با منوهای نرم افزار SPSS

گزینه های منوی File

با فرمان Save از منوی File می توانید تغییرات را روی فایل داده ها ذخیره کنید.

با فرمان Save As.. از منوی File می توان فایل داده ها را با نامی جدید (به غیر از نام فعلی فایلی که روی آن کار می کنید) ذخیره کرد.

با فرمان Save All Data می توانید تغییراتی را که به طور همزمان روی چند فایل داده فعال انجام داده اید، ذخیره کنید.

با فرمان Export To Database می توانید داده ها را به فرمتی دیگر برای استفاده در یک پایگاه داده دیگر صادر کنید. جزیئات دستور صدور فایل داده، به پایگاه داده دیگر، در بخش های بعدی توضیح داده شده است.

با فرمان Mark File Read Only می توان داده ها را به طور موقت به صورت فقط خواندنی درآورد تا نتوان تغییراتی در فایل داده ها ایجاد کرد و مجدداً از همین فرمان که اینک به صورت Mark File Read Write است، می‌توانید آنها را به حالت خواندنی نوشتنی درآورد.

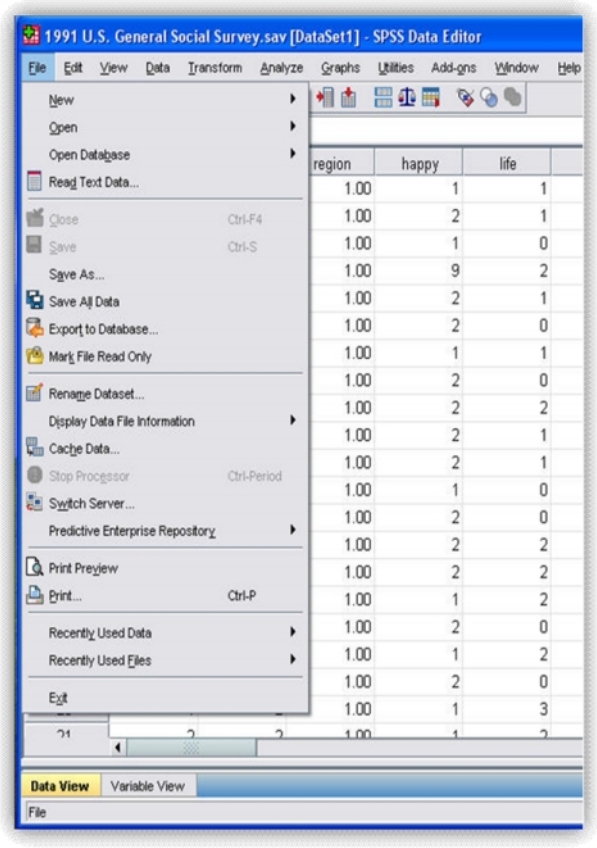
با فرمان Print Review می‌توان پیش نمایش نسخه چاپی از داده ها را مشاهده کرد.

با فرمان Print می‌توان داده‌ها را چاپ کرد. در این صورت قبلاً باید چاپگر شما روشن و آماده به کار باشد.

با فرمان Rename Dataset می‌توان نام پایگاه داده‌ها را تغییر داد.

با فرمان Recently Used Data تعدادی از فایل داده‌ها را که اخیرا توسط نرم افزار استفاده شده‌اند، نمایش می‌دهد. با این گزینه این امکان برای شما فراهم است تا با صرفه جویی در وقت، سریعتر به داده‌ها دسترسی پیدا کنید.

اگر فایل‌های دیگری غیر از فایل داده‌ها، مد نظر باشد، فرمان Recently Used File را به کار ببرید.

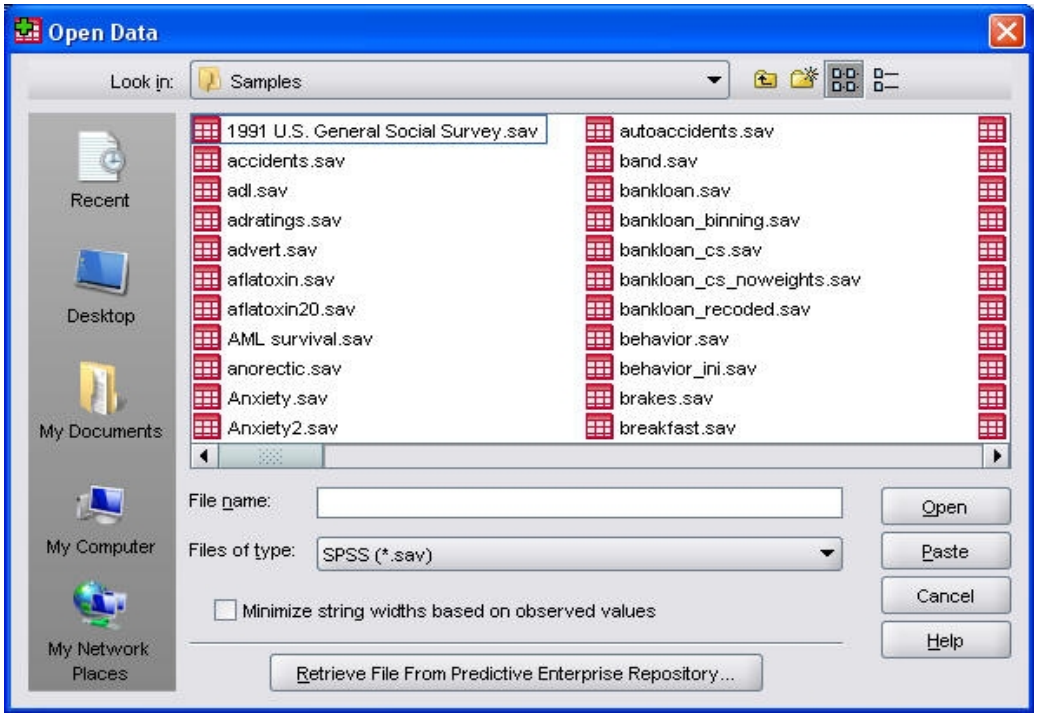


دسترسی به پوشه داده ها

هنگام نصب نرم افزار spss تعدادی فایل داده برای استفاده کاربران در پوشه‌ای به نام Sample قرار داده شده است. برای دسترسی به این داده‌ها اگر نرم افزار در پوشه Program File در درایو C نصب شده باشد، مسیر زیر را دنبال کنید:

C/Program File/SPSSLNC/SPSS17/Sample/

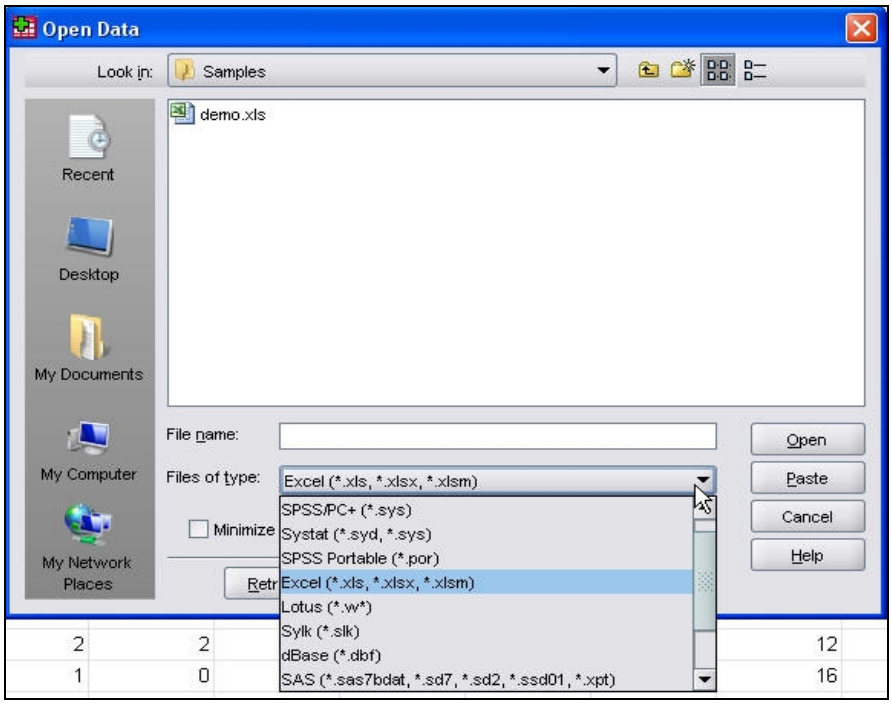
پس از باز شدن کادر محاور آن مانند شکل، فایل مورد نظر را انتخاب کنید و کلید Open را کلیک کنید.



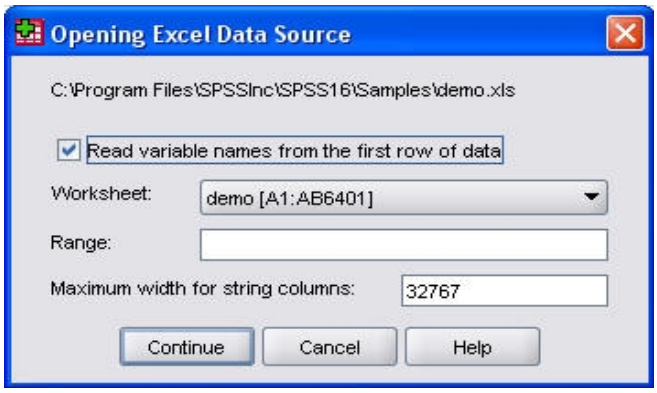
فراخوان فایل داده ها

در بعضی مواقع فایل داده هایی که برای مطالعه در اختیار دارید در یک نرم افزار صفحه گسترده مانند Excel یا Access تایپ شده اند و شما ناگزیرید آنها را به محیط SPSS وارد کنید. برای فراخوانی یک فایل صفحه گسترده مانند Excel مراحل زیر را دنبال کنید.

مسیر File/Open/Data را انتخاب کنید تا کادر محاوره ای مانند شکل باز شود.



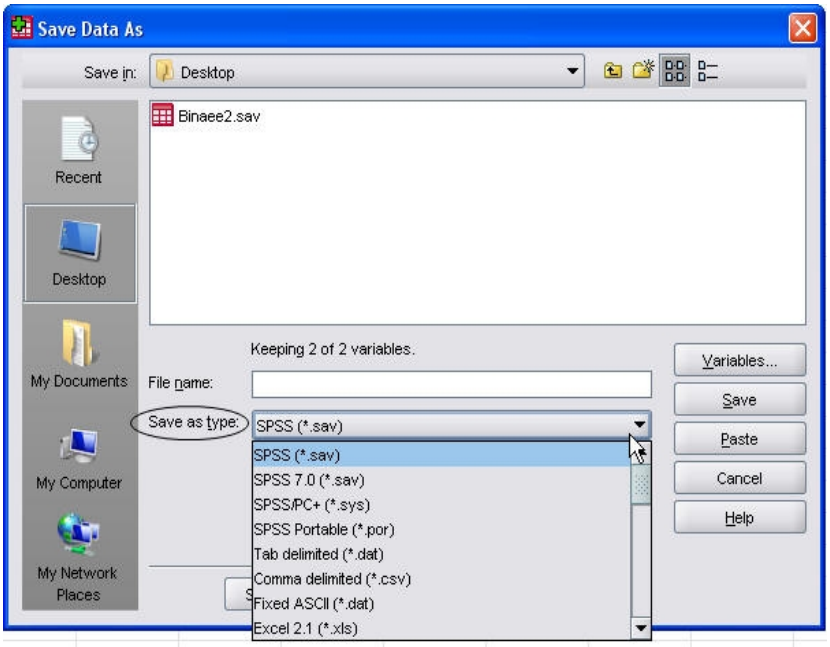
از نوار کرکره ای Files of Type فرمت Excel (*.xls) را انتخاب کنید.
از کرکره ای Look in: مسیری را که فایل اکسل در آن قرار دارد برای نرم افزار مشخص کنید.
نام فایل را در File Name: وارد کرده و کلید Open را کلیک کنید تا به کادر محاوره بعدی بروید.



در کادر محاوره باز شده گزینه Read Variable Name From the … را در صورتی که اسامی متغیرها در فایل صفحه گسترده نوشته شده است ، علامت دار کنید.
از کرکره ای Work Sheet: کاربرگی را که داده ها در آن وارد شده است برگزینید.
اگر قصد دارید محدوده خاصی از داده ها را به SPSS وارد کنید در بخش Range: و یا متغیرهای رشته ای را در کادر Maximum Width For… مشخص کنید. روی دکمه Open کلیک کنید تا داده ها به محیط SPSS وارد شوند.

ذخیره فایل داده ها

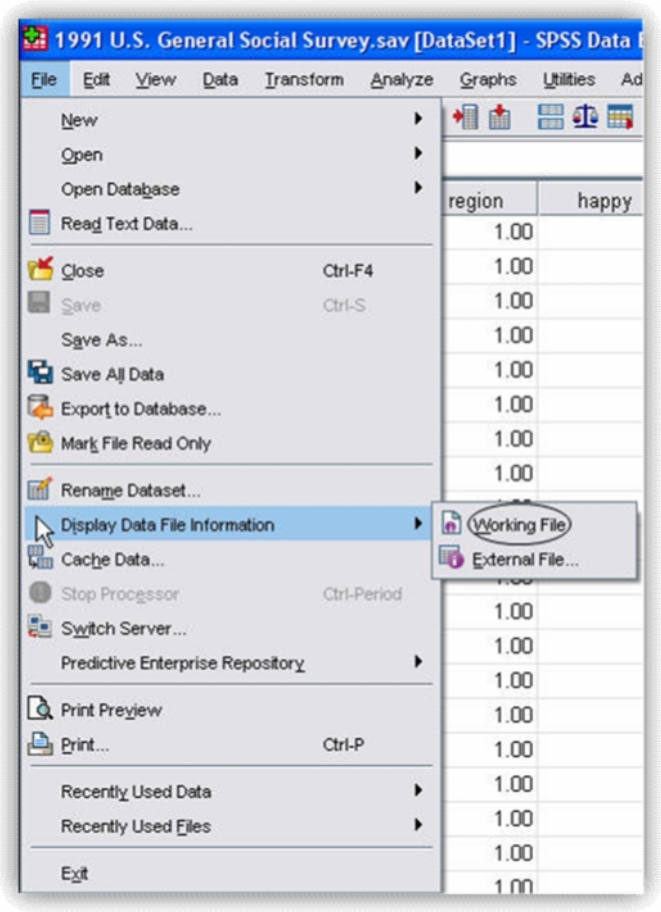
اگر قصد دارید داده‌های SPSS را به فرمتهای دیگری تبدیل کنید و از آنها در محیط نرم افزارهای دیگر استفاده کنید، مثلاً می‌خواهید یک فایل داده در SPSS را با فرمت Excel ذخیره نمائید، باید مراحل زیر را دنبال کنید:
فایلی را که می‌خواهید با فرمت دیگری غیر از SPSS ذخیره کنید، باز کنید.
فرمان File/Save as.. را اجرا کنید تا کادر محاوره ای مانند شکل بازشود.



از کشوی کرکره ای Save as Type: گزینه Excel 97 Through 2003 (*.xls) یـا Excel 2007(*.xls) را انتخاب کنید.
مسیری را که می خواهید فایل تبدیل شده را ذخیره کنید از Save in: معین کنید.
در بخش File name: برای فایل جدیدی که ایجاد می کنیم یک اسم انتخاب نمائید.

اطلاعات یک فایل داده

گاهی اوقات کاربران می‌خواهند به منظور خاصی اطلاعات مربوط بـه فایـل داده را در اختیـار داشـته باشـند، امکـان مشاهده جزئیات مربوط به متغیرها مانند: نام، نوع، مقادیر و برچسب‌های مقادیر آنها مفیـد مـی‌باشـد. مخصوصـاً اگـر فایل داده‌ای که کار می‌کنید بزرگ باشد و شما نام همه متغیرها را نتوانید به خاطر بسپارید.
شاید بخواهید اطلاعات مربوط به متغیرهای فایل داده‌ای را که در حال حاضر بر روی آن کار مـی‌کنیـد، در خروجـی مشاهده کنید و احتمالا از آن یک نسخه چاپی تهیه کنید.
در نوارمنو File گزینه Display Data File Information/Working File را انتخاب کنید.



نتیجه کار نام تمامی متغیرهای فایل حاضر همراه با مشخصات کاملی از آنها را در خروجی مشاهده خواهید کرد.

سایر منوهای نرم افزار:

منوی View

اگر می خواهید فونت، رنگ یا اندازه فونت مقادیر وارد شده در صفحه نمایش داده ها را تغییر دهید، از منوی View گزینه Font را انتخاب کنید و در کادر محاوره آن تغییرات دلخواه را اعمال کنید.

با فرمان Gridline از منوی View می توانید در پنجره نمایش داده ها، خطوط مرزبندی بین سلولها را مخفی یا آشکار کنید.

با فرمان Value Label از منوی View می توانید داده های اسمی و رتبه ای را در در صفحه نمایش داده ها، به یکی از دو صورت کد یا برچسب نمایش دهید.

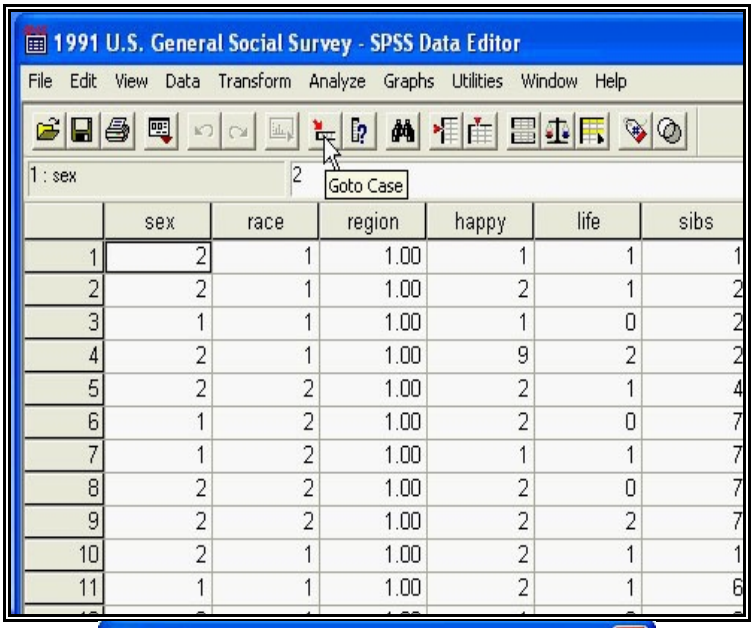
منوی Window

با فرمان Split از منوی Window می توان پنجره Data Editor را به چهار ناحیه تقسیم کرد و در صورت لزوم مرز نواحی با کلیک کردن و کشیدن موس تغییر داد. برای لغو فرمان فوق دوباره دستور Split در منوی Window را کلیک کنید. این کار برای فایلهایی که حجم وسیعی از داده ها را در بر دارند، مفید است.

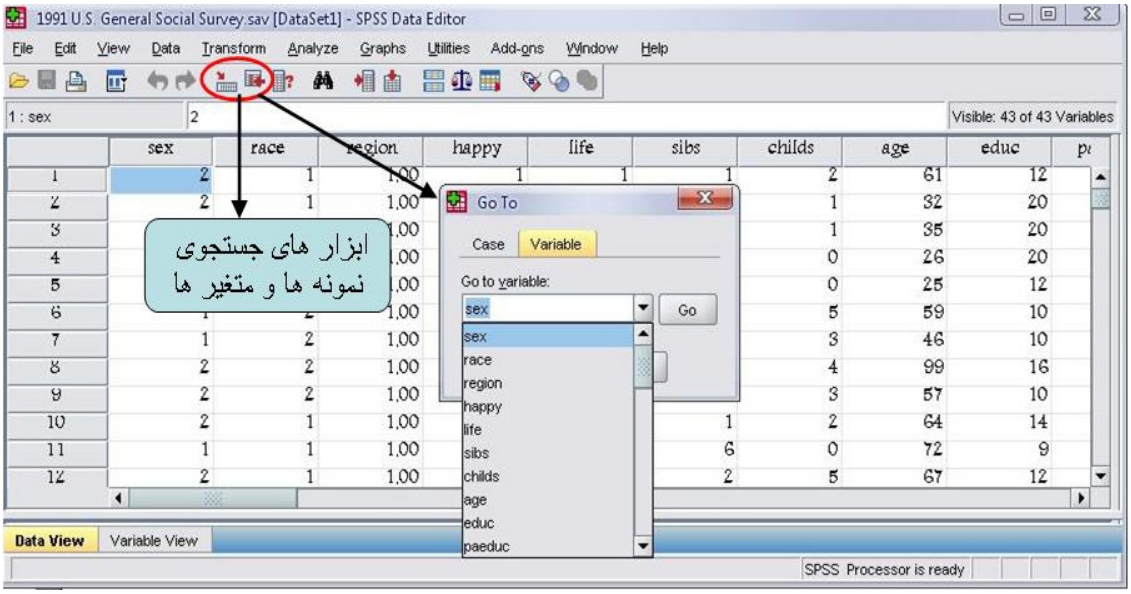
جستجو در فایل داده ها

منوی Edit

برای رفتن به یک نمونه (Case) مورد نظر از دستور Go To Case در منوی Edit و یا از نوار ابزار، آیکون Go To Case را انتخاب و د ر پنجره باز شده شماره Case مورد نظر را وارد کرده OK را فشار دهید.



همچنین برای رفتن به یک متغیر خاص میتوانید از گزینه Variable در منوی Edit و یا از نوار ابزار، آیکون Go To Variable را انتخاب کرده و نام متغیر مورد نظر را وارد کرده OK را فشار دهید.



برای رفتن به ابتدا و انتهای فایل داده ها به ترتیب از کلیدهای Ctrl+Home و Ctrl+ End استفاده کنید.

کلیه عملیات **حذف، اضافه، کات و کپی** کردن و .... یک متغیر یا یک نمونه (Case) یا بخشی از سلولها از گزینه های مناسب در منوی Edit و یا با راست کلیک کردن و انتخاب گزینه های مناسب امکان پذیر است.

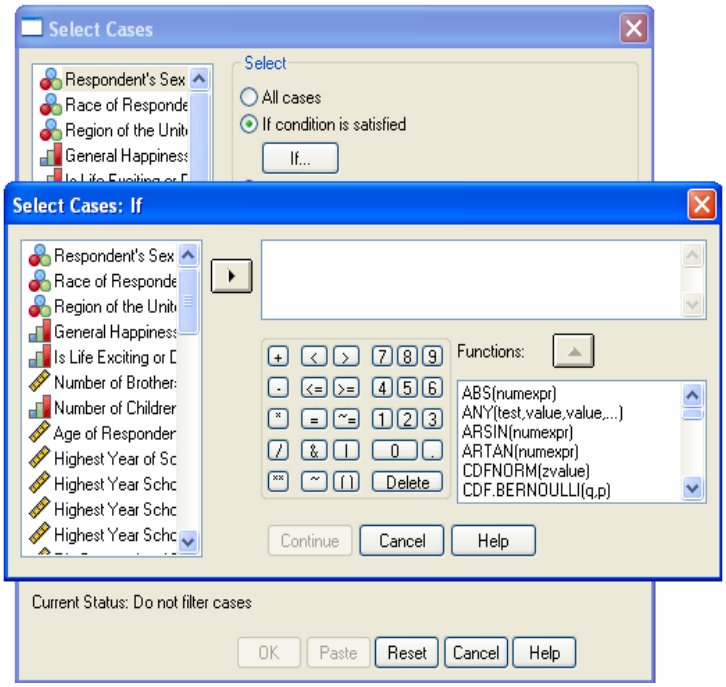
منوی Data

انتخاب نمونه ها (Select Case)

مواردی پیش می آید که بخواهیم تحلیل برای بعضی از نمونه‌های خاص انجام نشود. مثلا می‌خواهید کسانی را که زیر ۱۸ سال سن دارند به دلایلی از تحلیل کنار بگذارید و یا هنگامی که می‌خواهید یک نمونه تصادفی از بین نمونه‌ها می‌تواند از دستور select case استفاده کنید.

برای این کار می توانید از منوی اصلی گزینه Data و سپس Select Case را انتخاب کنید تا پنجره آن باز شود.

در گزینه اول (All Cases) همه داده ها در تحلیلها شرکت داده میشوند.



گزینه دوم If Condition is satisfied را انتخاب و برروی کلید if کلیک کنید تا کادر If باز شود از سمت چپ متغیر مورد نظرتان را انتخاب و به پنجره سمت راست منتقل کنید. سپس با استفاده از صحفه کلید داده شده دستورانتخاب را وارد کنید. برای مثال بالا : Age > =18

نتیجه کار به این صورت است که نمونه هایی که سن کمتر از ۱۸ سال دارند از مطالعه حذف خواهند شد.

گزینه سوم Random Sample Of Case برای انتخاب یک نمونه تصادفی در پنجره‌ی باز شده این گزینه را تیک بزنید. برای نحوه انتخاب کلید Sample را فشار دهید تا پنجره مربوط به آن باز شود. در پنجره باز شده:

الف اگر می‌خواهید درصدی از داده‌ها را بطور تصادفی انتخاب کنید مقدار آن را معلوم کنید.

ب تعداد نمونه مورد نظرتان را در گزینه بعدی وارد کنید و معلوم کنید که این تعداد از بین چه تعدادی از داده ها انتخاب شوند.

اگر می‌خواهید نمونه‌هایی را که شماره آنها بین دو مقدار مثلا بین ۵۰ تا ۱۵۰ است انتخاب کنید، این دو عدد را در جای خالی داده شده، وارد کنید. سپس Continue را فشار دهید.

اگر می‌خواهید نمونه‌ها را به دلخواه انتخاب کنید باید متغیری به عنوان فیلتر تشکیل دهید که شامل مقادیر صفر و غیر صفر باشد.

سپس باید در گزینه Use filter variable ابتدا متغیر فیلتر را وارد کرده و سپس ok را بفشارید. اگر متغیر فیلتر مقدار صفر باشد نمونه مورد نظر انتخاب نمی شود و اگر غیر صفر باشد نمونه مورد نظر انتخاب خواهد شد.

نتیجه کار برای همه گزینه های گفته شده در فایل داده ها به این صورت است که روی شماره بعضی از نمونه ها (Case) خط خورده که مفهوم عدم انتخاب و روی بعضی از شماره نمونه ها خط نخورده است که به مفهوم انتخاب شده است.

اگر می خواهید بجای خط خوردن نمونه های انتخاب نشده آنها را حذف کنید گزینه Delete را علامت دار کنید.

جزییات سایر گزینه های منوی Data به عهده خواننده گذاشته می شود.

منوی Transform

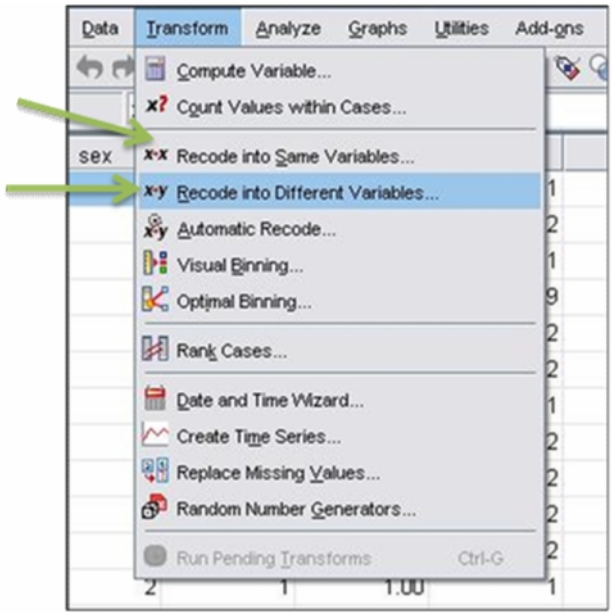
تغییر در داده ها (Recode)

گاهی اوقات می‌خواهیم داده‌های پیوسته را به چند طبقه دسته بندی کنیم (منظور تشکیل جدول توزیع فراوانی است) می‌توانید از دستورRecode استفاده کنید.

ابتدا از منوی اصلی به گزینه Transform بروید. دو گزینه پیش رو خواهید داشت:

1. Recode into Same Variables…

مقادیر تغییر یافته در همان متغیر ثبت شوند.

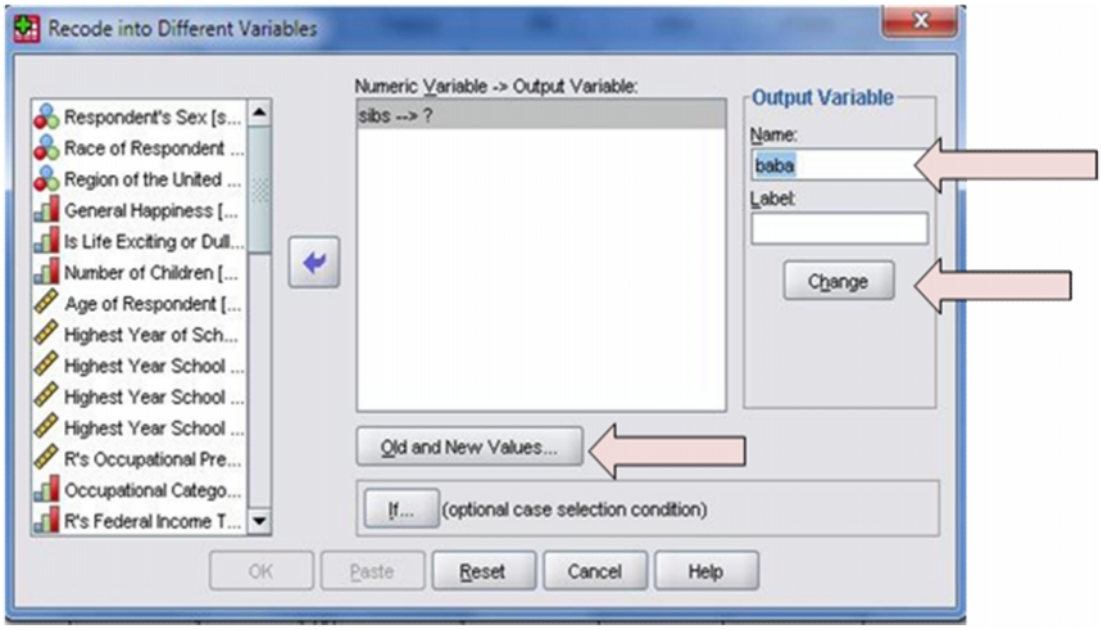


2-Recode into Different Variables…

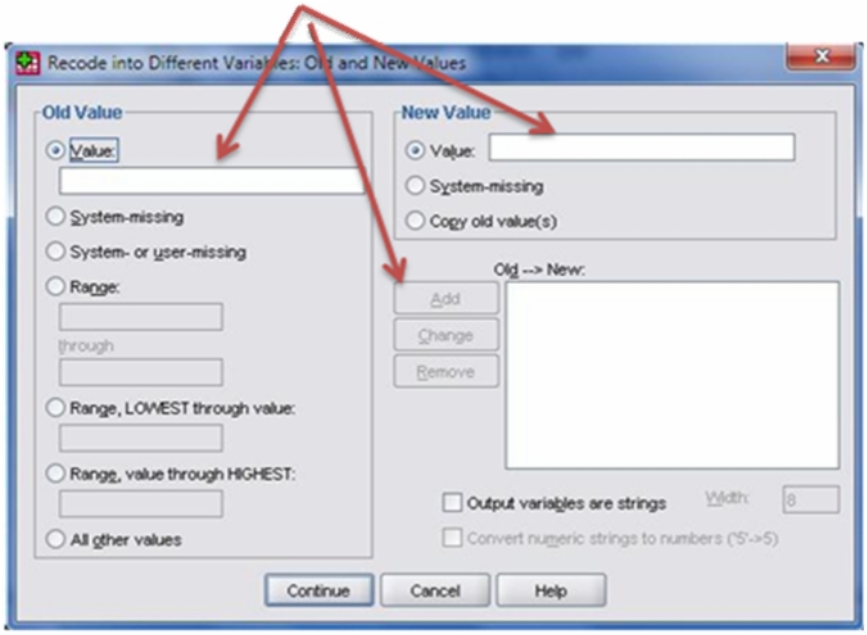
مقادیر تغییر یافته به متغیر جدید منتقل شوند.

در اینجا گزینه دوم، گزینه مناسب است. آن را انتخاب کنید تا به کادر محاوره مربوطه منتقل شوید.

ابتدا متغیری را که می‌خواهید تغییر دهید به کادر خالی سمت راست منتقل کنید. پس از انتقال، جلوی نام متغیر علامت سوال خواهد داشت که باید در کادر سمت راست(۱) ، یک نام وارد کنید و کلید Change را بزنید(۲) تا نام وارد شده، به جای علامت سوال قرار گیرد.



سپس گزینه Old and New Values را انتخاب کنید (۳) تا کادر محاوره مربوط به آن باز شود.



در این کادر محاوره چند امکان وجود دارد:

۱. اگر قصد تغییر یک یا چند مقدار از یک متغیر را دارید گزینه Value را انتخاب کنید که نحوه استفاده از آن به صورت زیر است:

مقدار متغیری را که می‌خواهید تغییر دهید در قسمت Old Value و مقدار جدید آن را در قسمت New Value وارد کنید. سپس کلید Add را بفشارید تا تغییرات به پنجره پایین منتقل شود. به همین منوال همه مقادیری که می‌خواهید تغییر دهید را اضافه کنید.

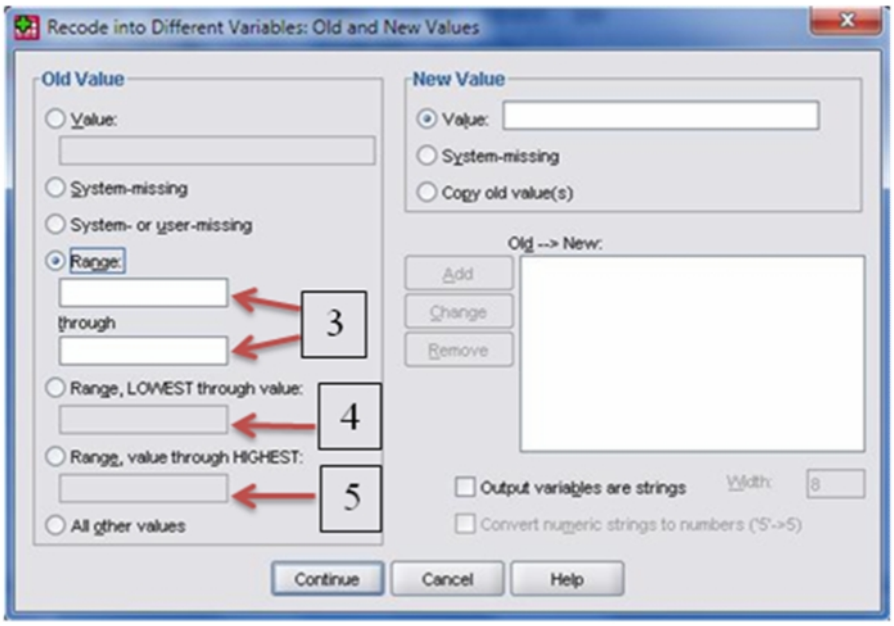
۲. برای تغییر مقادیر گمشده از گزینه: System missing استفاده کنید.

و برای نگهداشتن یک کپی از مقادیر قدیم از گزینه Copy Old Value(s) استفاده کنید:

۳. اگر مقادیر متغیر پیوسته‌ای را می‌خواهید تغییر دهید، گزینه Range را انتخاب کنید و حدود تغییرات را به شکل دو مقدار ابتدایی و انتهایی وارد کنید.

۴. در گزینه Range Lowest مقادیر کمتر از یک مقدار را وارد کنید.

۵. در گزینه Range Highest مقادیر بیشتر از یک مقدار را وارد کنید.



گر تغییرات را اعمال کرده‌اید و می‌خواهید برای باقیمانده مقادیر، مقدار جدیدی در نظر بگیرید، گزینه All Other Values را علامت دار کنید و برای آن مقداری تعیین نمایید.

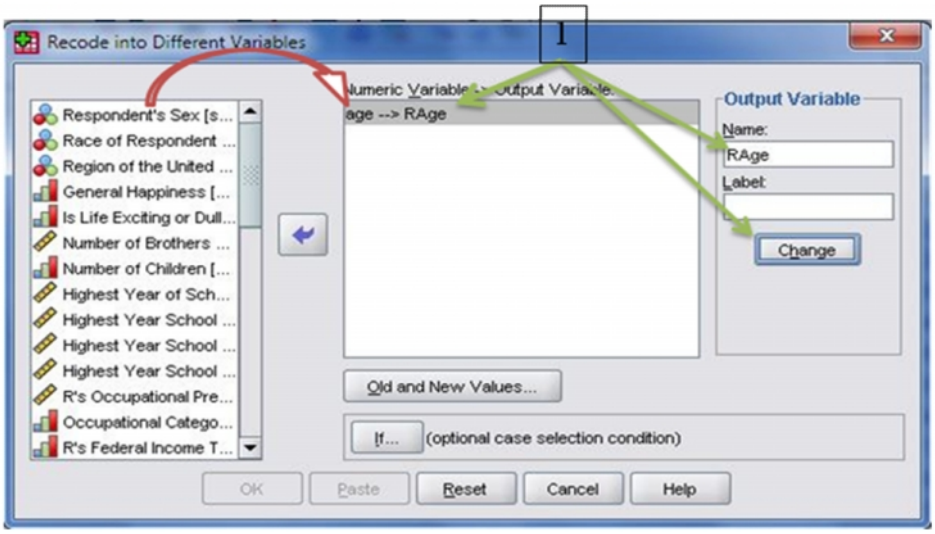
در پایان با انتخاب Ok تغییرات خواسته شده اعمال می شود و یک متغیر جدید به فهرست متغیر ها اضافه می‌گردد.

تمرین:

در مجموعهای داده ، با بررسی متغیر سن، در خواهید یافت که کمترین سن ۱۸ و بیشترین سن ۸۹ سال ثبت شده است. می‌خواهیم این متغیر را به ۷ سطح به صورت زیر طبقه بندی کنیم و جدول توزیع فراوانی را کامل نماییم.

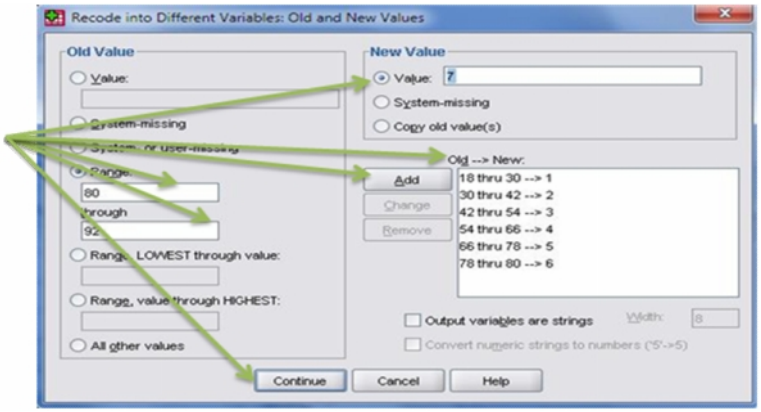
فراوانی طبقه	محدوده طبقه
؟	۱۸ – ۳۰
؟	۳۰ – ٤۲
؟	٤۲ – ٥٤
؟	٥٤ – ٦٦
؟	٦٦ – ۷۸
؟	۷۸ – ۸۰
؟	۸۰ -۹۲

برای انجام این کار، از نوار منو Transform گزینه Recode into Different Variable را انتخاب کنید تا پنجره Recode باز شود...



متغیر سن را به سمت راست منتقل کنید و در کادر Name، نام جدید مثلاً RAge را بـرای متغییـر جدیـدی کـه می‌خواهید مقادیر تغییر یافته در آن ثبت شود، تعیین کنید و گزینه Change را کلیک کنید.

برای اعمال تغییرات گزینه Old and New values را انتخاب کنید تا به کادر محاوره دیگری وارد شوید...



در کادر محاوره باز شده، گزینه Range را انتخاب کنید و در این قسمت اولین فاصله را وارد کنید.

با وارد کردن اولین فاصله (یعنی ۱۸ تا ۳۰)، در قسمت Value نیز مقدار ۱ را وارد کنید و دکمـه Add را نیـز بـه عنوان تایید، کلیک نمایید تا تغییرات به کادر :Old→ New اضافه شود.

برای بقیه فاصله ها به همین صورت ادامه دهید.

در پایان کلید Continue و Ok را به ترتیب کلیک کنید.

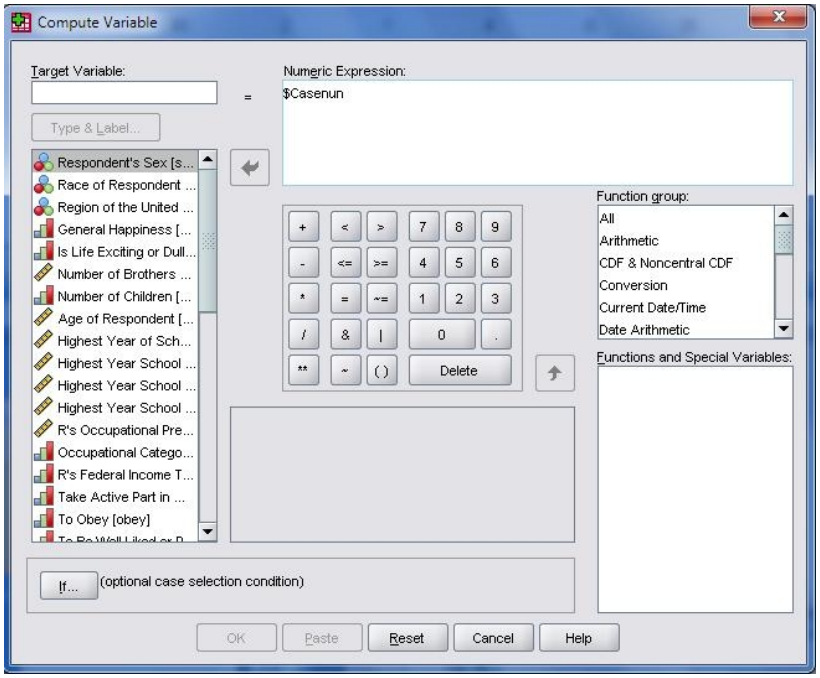
دستور Compute

دستور Compute برای ساختن متغیر جدید استفاده می گردد.

مثلاً : ساختن متغیر ردیف

معمولا در یک مجموعه داده، داشتن یک ستون که ردیف هر مورد را نمـایش دهـد، الزامـی بـه نظـر میرسـد. بـرای ساختن چنین ستونی باید از دستور Compute استفاده کرد.

ازنوار منو گزینه Transform را انتخاب کنید و سپس گزینه compute را کلیک کنید تا پنجره آن باز شود.



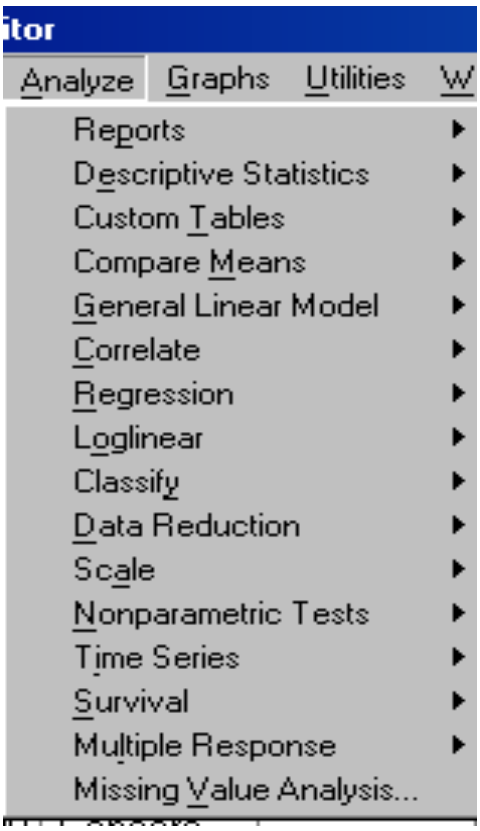
در مستطیل target variable یک اسم دلخواه وارد کنید. سپس در مستطیل Numeric Expression عبارت $Casenum را تایپ کرده و سپس OK را بزنید.

اینک یک ستون از شماره ۱ تا تعداد نمونه هایی که دارید در فهرست متغیرها اضافه شده است.

توجه کنید که نمی توان برای فایل داده خالی، شماره ردیف ایجاد کرد.

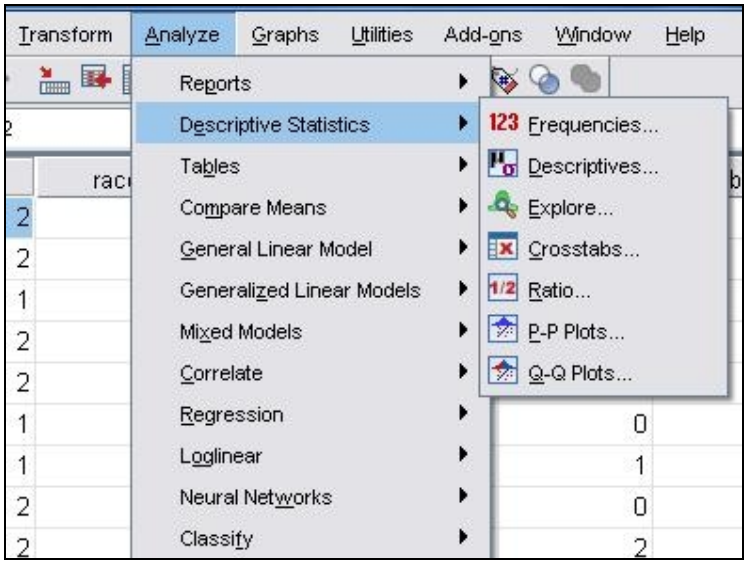
تحلیل داده ها

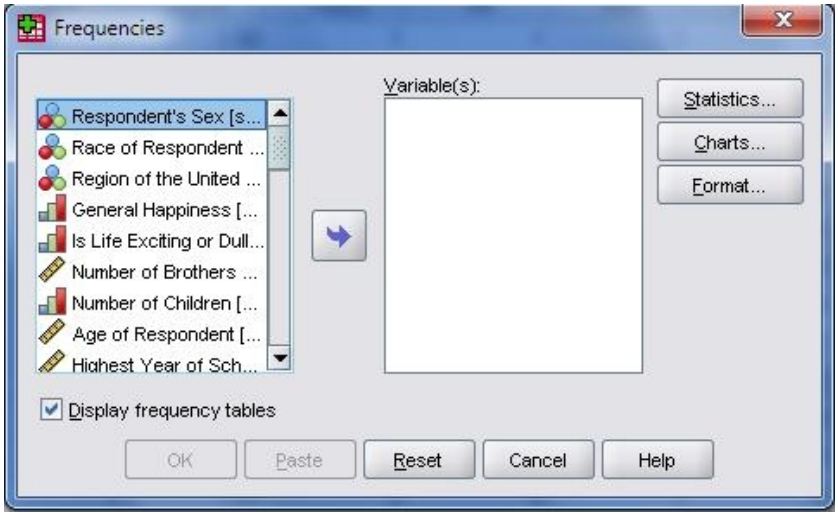
تحلیل داده ها با استفاده از **دستورات** (گزینه های) **نوار منو** Analysis انجام می گیرد.
تمامی گزارشهای آماری در مورد داده ها از جمله شاخصهای توصیفی شامل میانگین، انحراف معیار و ... جداول توصیفی و آزمونهای آماری شامل ضریب همبستگی، رگرسیون چند متغیره، و ... از طریق این منو قابل اجرا است.



جدول توزیع فراوانی (Frequencies)

در SPSS دستور Frequencies برای تهیه جدول توزیع فراوانی به همراه شاخص های مرکزی، چندکها، شاخص‌های پراکندگی و شاخص‌های توزیع، استفاده می گردد.
همچنین می‌توانید نمودارهای میله‌ای، دایره‌ای وهیستوگرام فراوانی را برای داده‌ها با انتخاب گزینه Charts در پنجره Frequencies ترسیم کرده و در صورت نیاز آن را با منحنی توزیع نرمال مقایسه کرد.
برای این منظور در نوار منو Analyze گزینه Descriptive Statistics و سپس گزینه Frequencies را انتخاب کنید تا پنجره مربوط به آن باز شود.





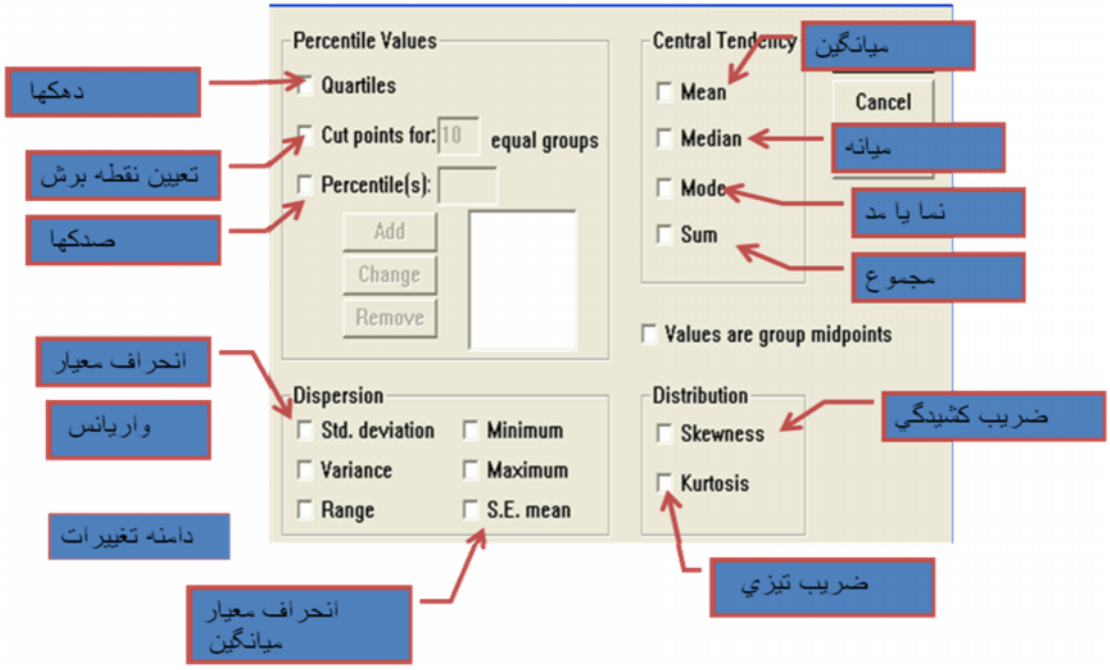
در این پنجره فهرستی از متغیرها در سمت چپ مشاهده می شود که با انتخاب یک یا چند متغیر از یک نوع و انتقال آن به سمت راست می توان برای هریک از آنها بطور همزمان جدول توزیع فراوانی رسم کرد.

توجه داشته باشید که گزینه Display frequency tables تیک خورده باشد.

در این حالت شما فقط یک جدول توزیع فراوانی ساده خواهید داشت.

توجه: جدول توزیع فراوانی برای متغیرهای کمی پیوسته مناسب نیست زیرا مقادیر زیادی از یک متغیر در یک جدول طولانی نمایش داده میشود. مگر آنها را قبلا دسته بندی کرده باشید.

اگر می‌خواهید آماره های توصیفی برای متغیرها محاسبه کنید گزینه Statistics را انتخاب و در پنجره مربوطه گزینه های دلخواه را علامت بزنید.



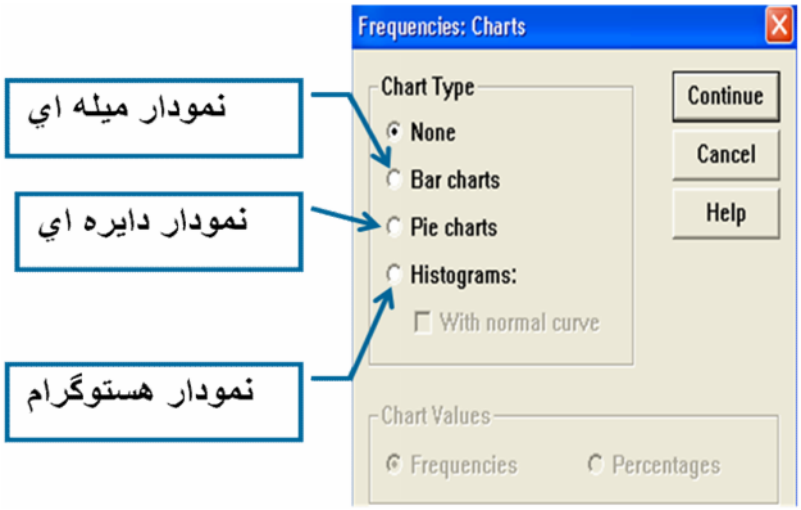
از بخش Percentile Value می‌توانید چارکها، دهک‌ها و صدک‌ها را محاسبه کنید.

از قسمت Central Tendency می‌توانید به ترتیب میانگین، میانه، نما و مجموع داده‌ها را محاسبه کنید.

از قسمت Dispersion می‌توانید انحراف استاندارد، واریانس، دامنه تغییرات، مینیمم، ماگزیمم و انحراف معیار و میانگین داده‌ها را به ترتیب محاسبه کنید.

در بخش Distribution قادر به محاسبه ضریب چولگی و ضریب کشیدگی خواهید بود.

اگر می‌خواهید ضمن داشتن جدول فراوانی نمودارهای آماری را رسم کنید، گزینه Charts را در پنجره Frequencies انتخاب نمایید تا پنجره Frequencies: Chart باز شود.



۱ با انتخاب گزینه Bar Cart یک نمودار ستونی برای داده‌های گسسته و کیفی رسم خواهد شد.

۲ اگر گزینه Pie Chart را انتخاب کنید یک نمودار دایره‌ای برای متغیرهای کیفی خواهید داشت.

۳ گزینه Histograms برای داده‌های پیوسته، یک نمودار مستطیلی رسم خواهد کرد.

۴ اگر می‌خواهید مقادیر نمودار برحسب فراوانی مطلق یا درصد فراوانی باشد، یکی از گزینه‌های Chart Values را انتخاب کنید.

شاخص های توصیفی (Descriptive)

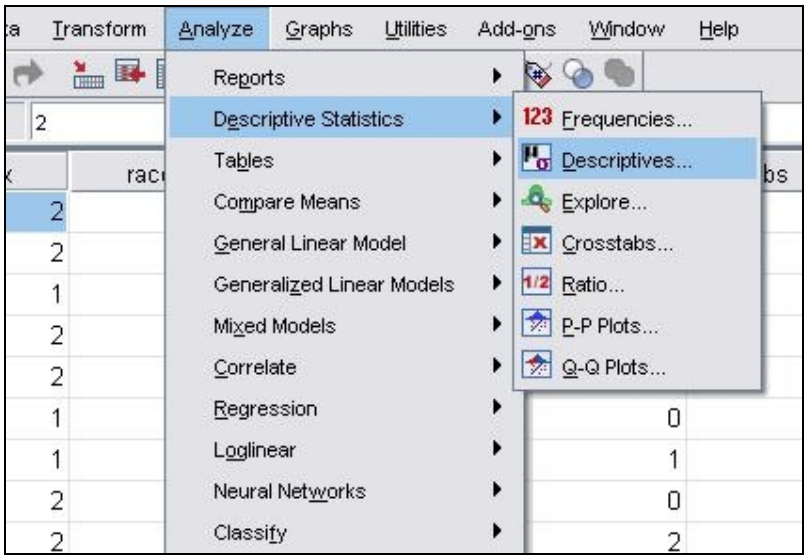
دستور Descriptive Statistics برای محاسبه بعضی از شاخص‌های توصیفی برای داده‌های کمی است. با این دستور شما می‌توانید شاخص‌های پراکندگی، شاخص‌های توزیع و میانگین و مجموع داده‌ها را محاسبه کنید.

از منوی اصلی Analyze و سپس گزینه Descriptive Statistics و بعد گزینه Descriptive را انتخاب کنید تا پنجره آن مانند شکل ظاهر شود.

Menu: ***Analyze***

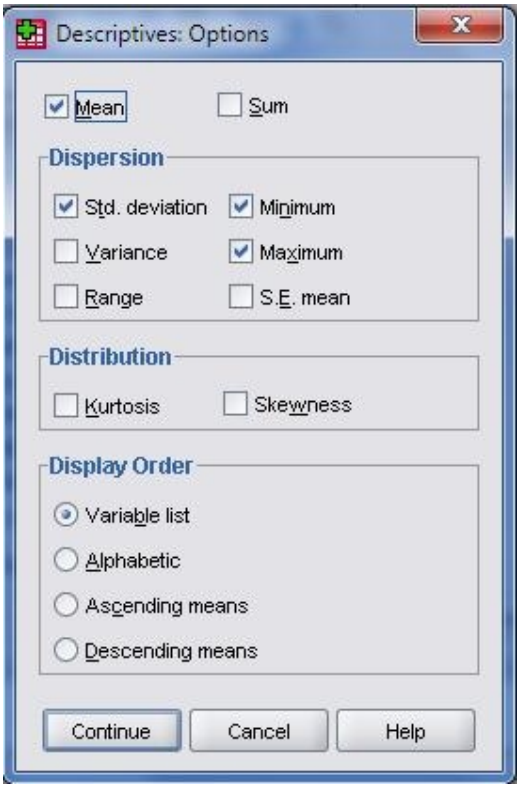
Descriptive Statistics…

Descriptives



در این پنجره فهرستی از متغیرها در سمت چپ مشاهده می شود (در این پنجره فقط متغیرهای کمی را مشاهده می‌کنید) که با انتخاب متغیرهای مورد نظر آنها را Highlight نموده با دکمه جهت دار آنها را به کادر روبرو انتقال می دهیم. در انتها گزینه OK را کلیک می کنیم.

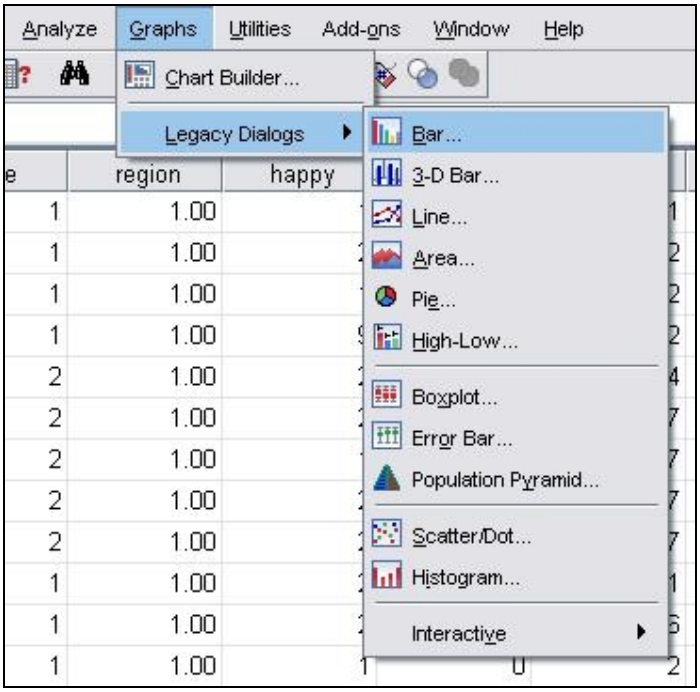
با استفاده از گزینه Options می‌توان برای هریک از آنها به طور همزمان شاخص‌های توصیفی را نیز محاسبه نمود.
با کلیک بر روی Options پنجره Descriptive Options باز می‌شود.



۱ در این پنجره با انتخاب دو گزینه Mean و Sum می توانید میانگین و مجموع داده ها را بدست آورید.
۲ از قسمت Dispersion می‌توانید با تیک زدن گزینه‌های مناسب انحراف استاندارد، واریانس، دامنه تغییرات، مینیمم، ماگزیمم و انحراف معیار و میانگین داده ها را به ترتیب محاسبه کنید.
۳ در بخش Distribution قادر به محاسبه ضریب چولگی و ضریب کشیدگی توزیع داده ها خواهید بود.
۴ در قسمت Display Order می‌توانید خروجی‌ها را با ترتیب‌های داده شده ببینید.
۵ با کلیک بر روی continue و ok می‌توانید حاصل عملیات را در خروجی مشاهده کنید.

نمودارهای آماری

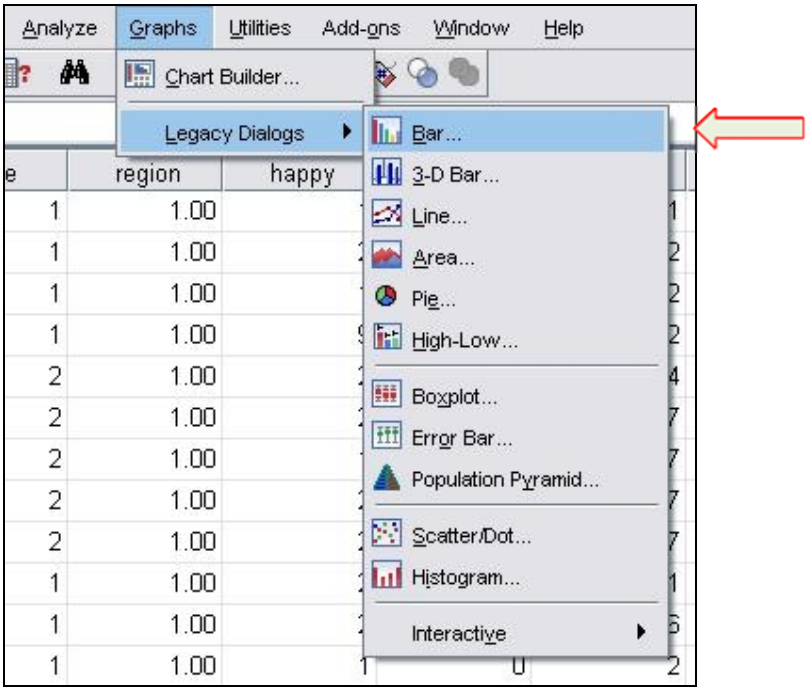
جهت رسم انواع نمودارها از منو Graphs استفاده می شود. نمودارهای میله ای، ستونی، خطی، دایره ای، پراکنش و....
SPSS نمودارهای متعددی برحسب نوع متغیر ها، در دو دسته کلی در اختیار شما می‌گذارد.
۱ نمودارهای Interactive (اثر متقابل) که حداقل در آن دو متغیر شرکت دارند و برای نمایش اثرات متقابل متغیر ها از این نمودارها استفاده می‌کنیم.
۲ نمودارهای توصیفی ساده که برای نمایش شکل توزیع فراوانی از آنها استفاده می‌کنیم.
اگر می‌خواهید یک نمودار ساده برای داده ها رسم کنید از منوی اصلی گزینه Graphs و گزینه Legacy Dialogs را انتخاب و سپس نمودار دلخواه را انتخاب کنید.



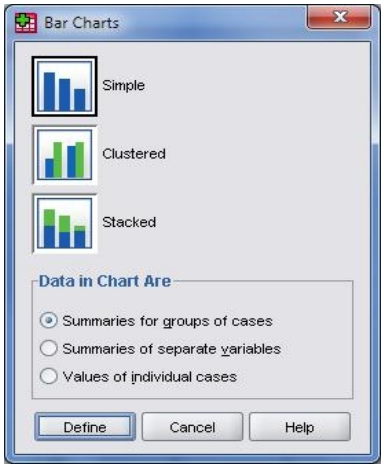
نمودارهای ستونی

نمودار ستونی (Bar charts) این نمودار برای متغیرهای پیوسته که طیف وسیعی از مقادیر را در بر می گیرند مناسب نیست ولی در بسیاری موارد که متغیرها دارای سطوح کمتری هستند بکار می‌رود.

اگر می‌خواهید یک نمودار ستونی ساده برای متغیر دلخواه ترسیم کنید. ابتدا از منوی اصلی گزینه Graphs را انتخاب کنید و گزینه Bar را کلیک کنید تا کادر محاوره Bar Chart باز شود.



در هر یک از نمودارهای ستونی به طور پیش فرض گزینه N of Case که تعداد نمونه را به عنوان فراوانی ستونها در نظر می‌گیرد، انتخاب شده است.



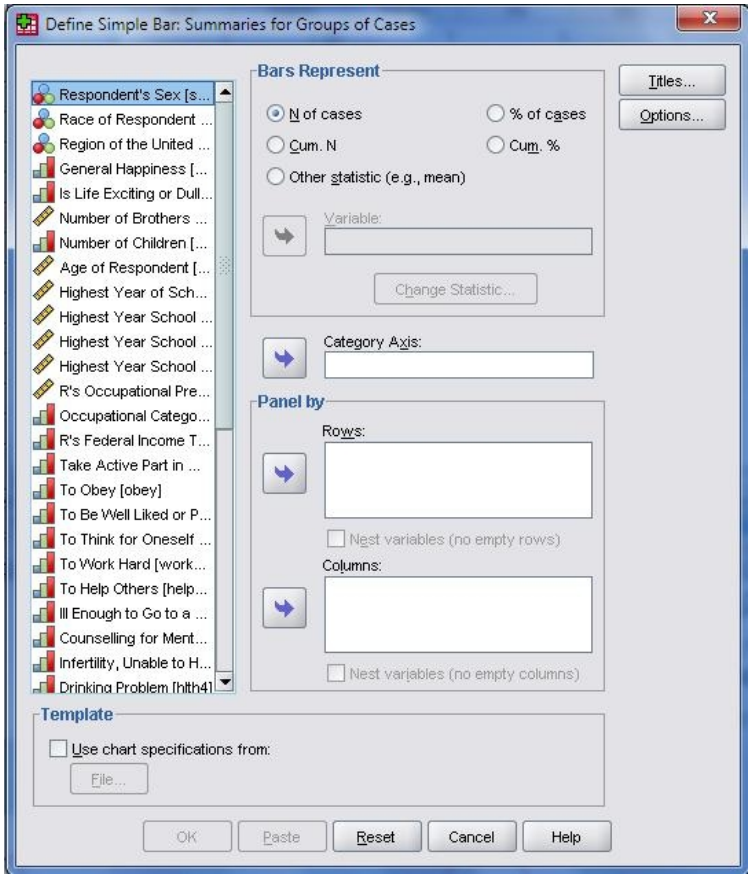
اگر می‌خواهید ستونها، فراوانی درصدی باشند، گزینه % of case را انتخاب کنید.

اگر می‌خواهید ستونها، فراوانی تجمعی باشند گزینه cum.n of case را انتخاب کنید.

اگر می‌خواهید ستونها، فراوانی تجمعی نسبی باشند گزینه cum.% of case را انتخاب کنید.

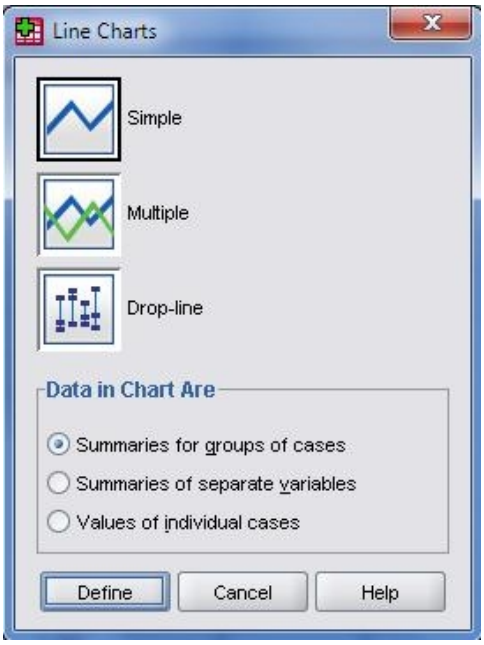
اگر می‌خواهید ستونها، میانگین متغیر دیگری باشند گزینه other summary function را انتخاب و در پنجره آن متغیر مورد نظرتان را وارد کنید.

اگر به جای میانگین متغیر دیگر، می‌خواهید آماره دیگری (مانند میانه و واریانس و ...) انتخاب کنید گزینه change statistics... را فشار دهید و در پنجره آن یک گزینه را انتخاب کنید.



نمودار خطی

نمودار خطی (line chart) این نمودار به سه شکل زیر وجود دارد:



۱ نمودار ساده (simple).

۲ چند گانه (multiple) که برای نمایش توزیع دو متغیر نسبت به هم استفاده می شود.

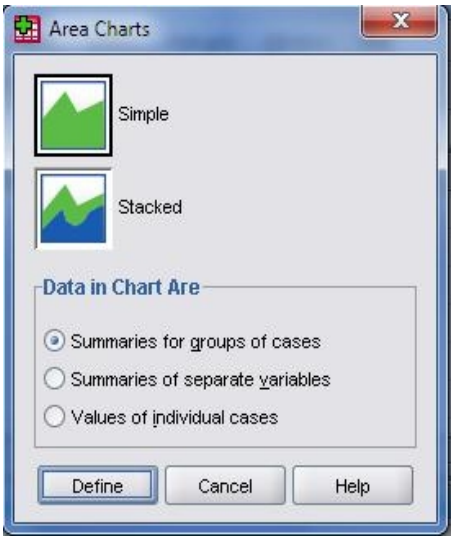
۳ تکه خطی (drop - line) که کمینه و بیشینه دو متغیر نسبت به هم را نمایش میدهد.

ابتدا در بخش Data in chart Are محور نمودار را انتخاب نمایید.

با انتخاب گزینه Define مانند قبل متغیر مورد نظر وگزینه های مناسب را انتخاب کنید تا یک نمودار خطی رسم شود.

نمودار سطحی

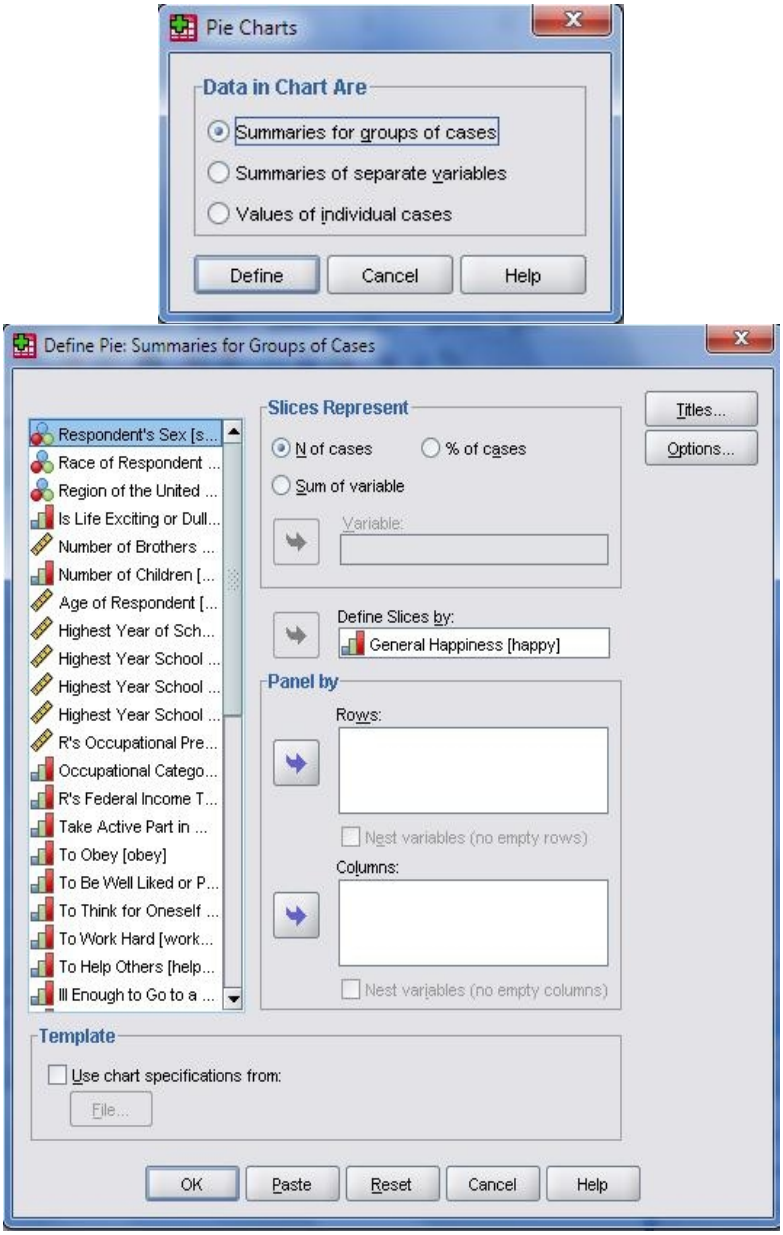
نمودار سطحی (area chart) مانند نمودار خطی است که بجای خطوط نواحی مربوط به متغیر نمایش داده می شود.



از منوی اصلی گزینه Graphs را انتخاب کنید و گزینه Area را کلیک کنید تا پنجره Area Chart باز شود.
در این پنجره اگر گزینه Simple را انتخاب کنید می‌توانید برای یک متغیر نمودار سطحی رسم کنید.
اگر گزینه Stacked را انتخاب کنید می‌توانید برای دو متغیر و به صورت پشته ای این نمودار را رسم کنید.
با انتخاب گزینه Define و انتخاب متغیر مورد نظر می‌توانید نمودار سطحی رسم کنید.

نمودار دایره ای

نمودار دایره‌ای (pie chart) از این نمودار برای نمایش وضعیت متغیرهای کیفی استفاده می‌کنیم هر چند در مواردی برای متغیرهای دیگر نیز استفاده می‌شود.
از منوی اصلی گزینه Graphs را انتخاب کنید و گزینه Pie را کلیک کنید تا پنجره Pie Chart (شکل سمت راست) باز شود. کافی است کلید Define را کلیک کنید تا کادر محاوره سمت چپ باز شود.



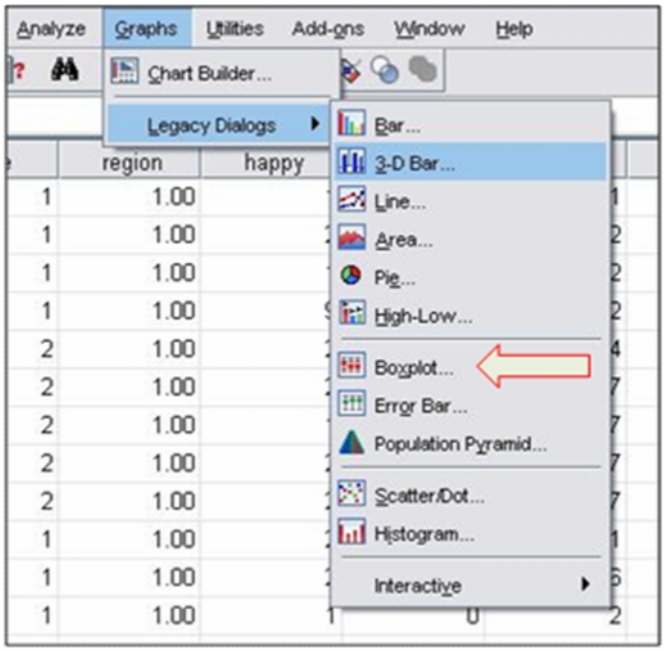
در این کادر محاوره متغیر مورد نظر را به کادر Defin Slices By: ببرید و گزینـه مناسـب را انتخـاب کنیـد. در ادامه کلید OK را انتخاب کنید تا نمودار دایره ای رسم شود.

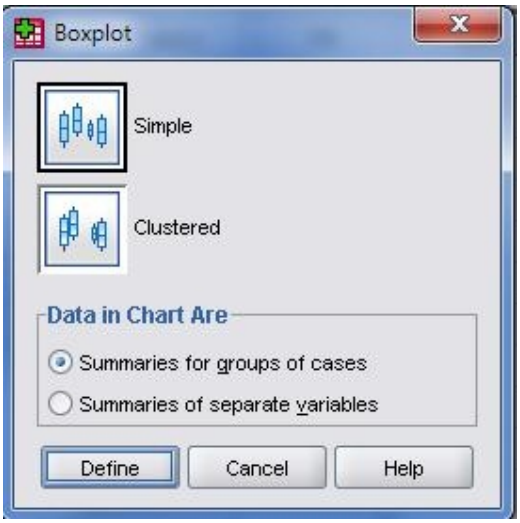
نمودار جعبه ای

نمودار جعبه ای (Box plot) – این نمودار برای نمایش وضعیت متغیرهای کمی بسیار مناسب است زیـرا در رسـم آن از آماره های میانه، چارکها و مقادیر کمینه و بیشینه یک متغیر استفاده می شود.

این نمودار به صورت ساده (simple) برای نمایش وضعیت یک متغیر و بـه صـورت خوشـه ای (clustered) بـرای مقایسه وضعیت دو متغیر در برابر هم استفاده میشود.

از منوی اصلی گزینه Graphs را انتخاب کنید و گزینه Box plot را کلیک کنید تا پنجره مربوط به آن باز شود.



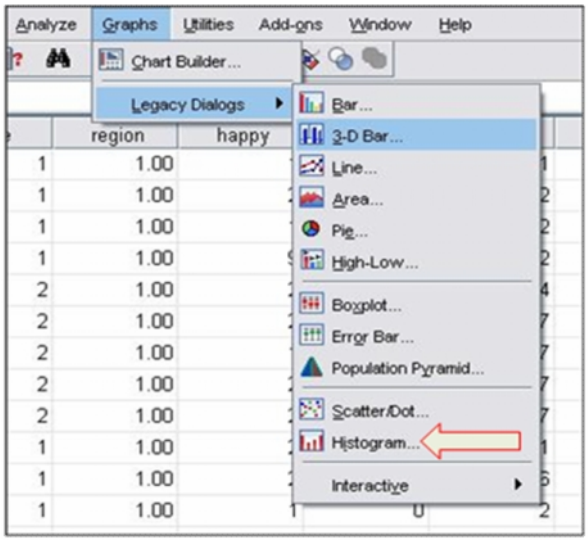


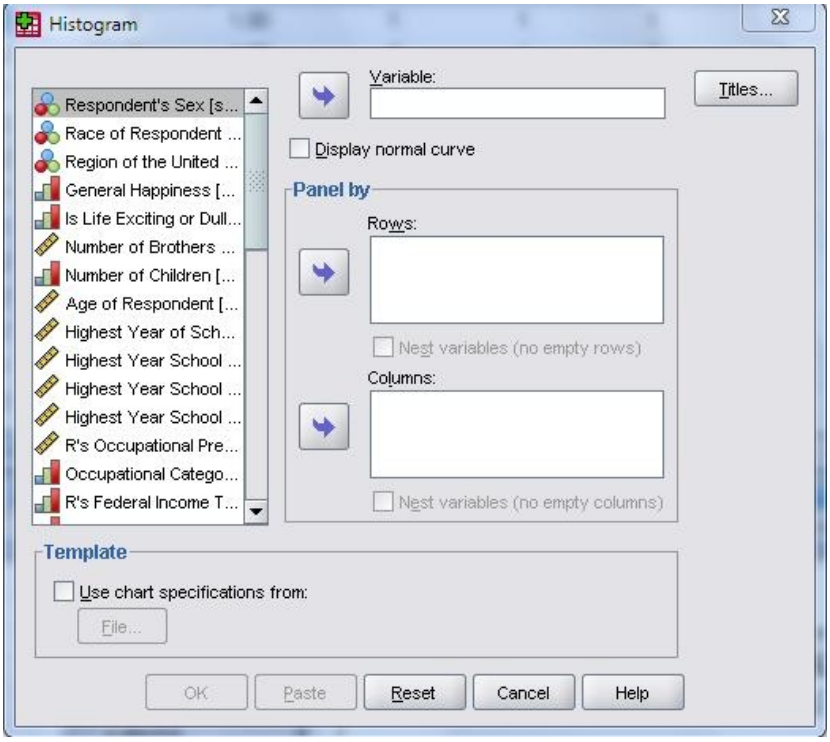
نمودار جعبه ای در این پنجره با انتخاب گزینه Simple می‌توانید یک نمودار جعبه ای برای یک متغیر رسم کنید.

اگر گزینه Clustered را انتخاب کنید می‌توانید یک نمودار به صورت خوشه‌ای برای دو یا چند متغیر رسم کنید.
بر روی گزینه Define کلیک کنید و سپس با انتخاب متغیرهای مناسب یک نمودار جعبه ای رسم کنید.

نمودار مستطیلی

نمودار مستطیلی (Histogram)- از این نمودار برای نمایش توزیع متغیرهای کمی استفاده کنید و در نظر داشته باشید که برای متغیرهای پیوسته که طیف وسیعی از مقادیر را در بر میگیرند مناسب نیست.
از منوی اصلی گزینه Graphs را انتخاب کنید و گزینه Histogram را کلیک کنید تا پنجره مربوطه باز شود.





متغیر مورد نظر به راست پنجره منتقل کنید از گزینه Display Normal Curve برای مقایسه توزیع نمودار داده ها با توزیع نرمال استفاده کنید.
سپس بر روی OK کلیک کنید تا نمودار مستطیلی مورد نظرتان رسم کنید.

نمودارهای ساقه و برگ

نمودار ساقه و برگ Stem & leaf این نمودار نحوه توزیع داده‌ها را به خوبی وصف می‌کند به این صورت که شامل تعدادی ساقه و تعدادی برگ است.

```
Age of Respondent Stem-and-Leaf Plot for
SEX= Male

 Frequency    Stem &  Leaf

     3.00        1 .  9&
    72.00        2 .  000111111111222222222333333333444444
    65.00        2 .  555555555666666777777888888999999
    76.00        3 .  000000000111111222222222333333344444444
    82.00        3 .  555555555556666677777888888888999999999
    77.00        4 .  000000001111111222223333333334444444
    53.00        4 .  555566777777788888899999
    41.00        5 .  0001111222223333444
    27.00        5 .  5556677888999
    38.00        6 .  0001112222233333344
    27.00        6 .  5556777888999
    41.00        7 .  00001112222333344444
    17.00        7 .  5556789
    13.00        8 .  12223&
     4.00        8 .  9&

 Stem width:  10
 Each leaf:       2 case(s)
```

در پایین نمودار اندازه ساقه و تعداد مورد در هر برگ آمده است. به عنوان مثال چون اندازه ساقه ۱۰ است، در عدد ۲۳ ساقه آن ۲ و برگ آن ۳ در نظر گرفته‌شده است و همین‌طور چون هر برگ شامل ۲مورد است فراوانی‌ها دو برابر تعداد برگها است.
بنابراین این نمودار ساقه و برگ اطلاعات دقیقتری راجع به مقادیر یک متغیر را معلوم می‌کند.
برای اینکه ساقه و برگ را به مقادیر حقیقی تبدیل کنیم باید مقدار ساقه را در پهنای ساقه ضرب کنیم و سپس با برگ آن جمع کنیم.

دستور Explore

یکی از امکانات مفید برای توصیف اطلاعات و آنالیز اکتشافی داده‌ها، دستور Explore است که دارای امکانات فراوانی است.

- اگر می خواهید در بین داده ها، اطلاعات بیشتری از نمونه ها را جستجو کنید و موشکافانه تر به جزئیات داده ها نگاه کنید. دستور Explore را به عنوان یک دستور جامع آنالیز اکتشافی داده ها به کار بگیرید.

در طریق Explore می توانید:

مقادیر یک صفت کمی را در سطوح یک متغیر کیفی و با استفاده از نمودار یا شاخص‌های توصیفی مقایسه کرد. برای خلاصه کردن و توصیف مشاهدات از نمودارهای مناسب بهره گیری نمایید.

جدول‌های خلاصه شده‌ای از اطلاعات مربوط به نمونه ها و زیر گروههایی از نمونه ها را بدست آورید.

در میان اطلاعات، نمونه‌های غیرعادی را که مقادیر آنها از سایر نمونه‌ها بیشتر یا کمتر هستند، کاوش کنید.

شاخص‌های آماری مناسب را در گروهها محاسبه کرده و فاصله اطمینان برای میانگین به دست آورید .

می‌توانید نمودار ساقه و برگ و هیستوگرام را برای نمایش توزیع فراوانی در بین گروهها رسم کنید.

نمودار مفید جعبه ای را برای مقایسه شاخص‌های میانه، چارک اول و سوم بین گروهها بکار گیرید.

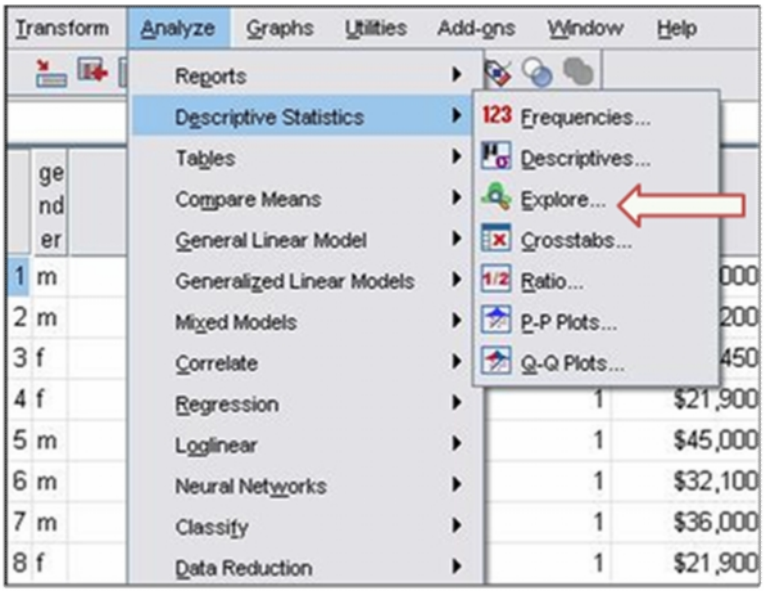
می توانید از نمودار Q-Q برای بررسی نرمال بودن مقادیر متغیر وابسته در هر گروه استفاده کنید.

همچنین می‌توان از این طریق برای مقایسه گروهها نیز استفاده کرد.

از این طریق متغیر کمی (Scale) را به عنوان متغیر وابسته و متغیر کیفی (Nominal , Ordinal) را به عنوان متغیرمستقل در نظر می‌گیریم.

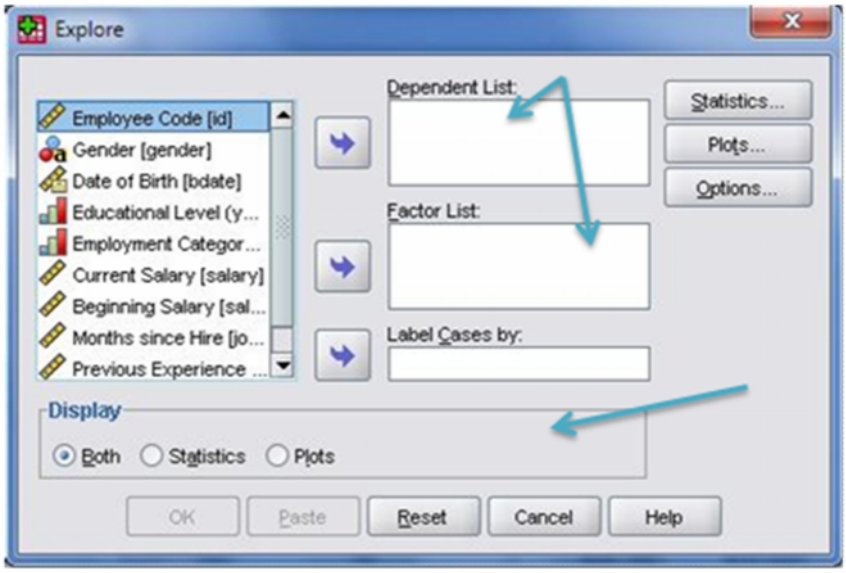
می توانید از نمودار Q-Q برای آزمون نرمال بودن متغیر وابسته در هر گروه استفاده کنید.

از منو اصلی گزینه Analyzed، گزینهdescriptive statistics و سپس explore را انتخاب کنید.



۱- به پنجره dependent list یک متغیر کمی مانند سن/ تعداد فرزند و ... و به پنجره factor list یک متغیر کیفی مثل جنس/ گروه خون و... منتقل کنید.

۲ اگر می‌خواهید فقط شاخص ها را محاسبه کنید، در پایین و سمت چپ پنجره، گزینه Statistics و اگر می‌خواهید فقط نمودار داده ها را مشاهده کنید گزینه plots و اگر می‌خواهید از هر دو مورد استفاده کنید گزینه Both را علامت دار کنید.



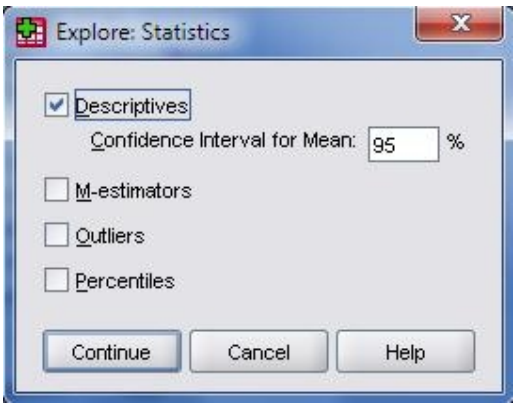
روی گزینه statistics کلیک کنید تا کادر محاوره آن (explore: statistics) باز شود. در این کادر محاوره می‌توانید هر یک از گزینه های زیر را انتخاب کنید:

بطور پیش فرض گزینه descriptive علامت دار شده است که منجر به محاسبه شاخص های توصیفی و همچنین یک فاصله اطمینان ۹۵٪ برای میانگین خواهد شد.

گزینه M-estimator، منجربه محاسبه آماره‌ای می‌شود که به هر داده بسته به فاصله آن از میانگین، وزن می‌دهد.

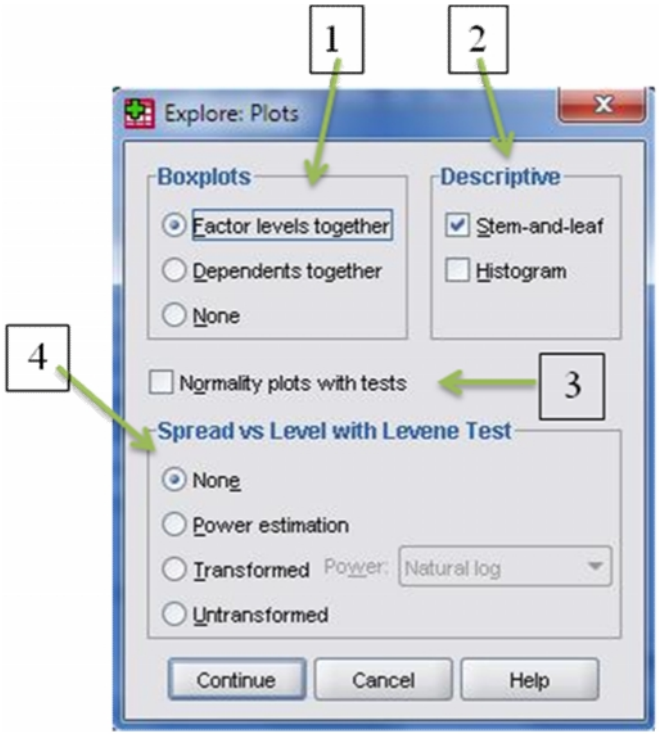
گزینه Out line پنج مورد از بزرگترین و کوچکترین مقادیر متغیر وابسته را نمایش می‌دهد.

گزینه percentiles صدکهای ضروری ۵و۱۰و۲۵و ۵۰ و۷۵ و ۹۰و ۹۵ را محاسبه می‌کند.



توضیح اینکه فاصله‌ی اطمینان مذکور، فاصله‌ای است که با اطمینان ۹۵ درصد، میانگین جامعه را در بر می‌گیرد.

.



- اگر گزینه plots را انتخاب کنید پنجره مربوط به explore: plots (مانند شکل) باز شده و در آن می‌توانید نمودارهای زیر را رسم کنید.

۱- نمودار جعبه ای با گزینه های :

الف سطوح متغیر فاکتور با هم

ب سطوح متغیر وابسته با هم

ج هیچکدام

۲- نمودار های توصیفی با دو گزینه:

الف رسم نمودار ساقه و برگ

ب رسم هیستوگرام فراوانی

- اگر گزینه Normality plots with tests را علامتدار کنید. آزمون نرمال بودن توزیع را می توانید با استفاده از **نمودار Q-Q** انجام دهید.
- گزینه Spread vs. level with levene test مربوط به آزمون لون برای یکسان بودن واریانس ها در سطوح هر یک از متغیرهای فاکتور است.

تمرین:

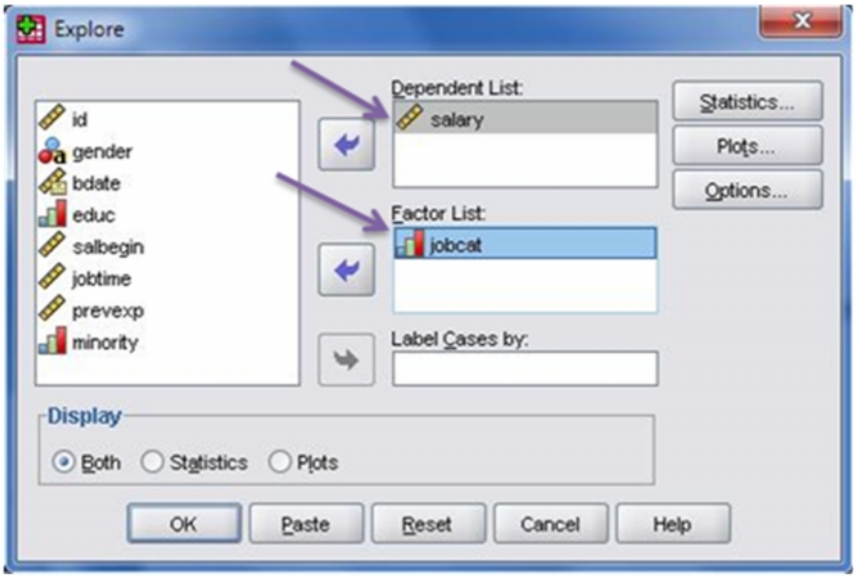
از مجموعه داده های spss فایل داده‌ی استخدامی (Employee) را معرفی می‌کنیم. در این فایل اطلاعاتی ۴۷۴ نمونه از بین شاغلین انتخاب شده و از هر نمونه ۱۰ متغیر اندازه گیری شده است.

فایل داده‌های Employee را از مسیر زیر بازخوانی کنید و برای مقایسه حقوق شاغلین (Salary) در هر گروه شغلی (Jabcat)، شاخص‌های توصیفی را بدست آورده نمودار مناسب رسم کنید.

Program file/ SPSSInc/ SPSS16/ Samples/ Employee.sav

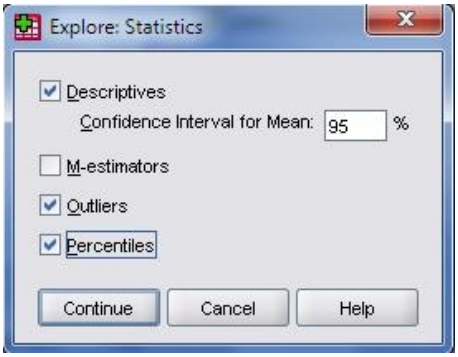
تعدادی از متغیرهای این فایل داده در جدول زیر آورده شده است.

نام متغیر	توضیح	ماهیت
gender educ jobcat salary jobtime	جنسیت (مرد- زن) تحصیلات (برحسب سال) گروه شغلی (اداری- نگهبانی- مدیریت) حقوق فعلی سابقه کار فعلی	Nominal Nominal Scale Scale Scale



در کادر محاوره Explore متغیر حقوق (Salary) را به کادر متغیرهای وابسته (Dependent list) و متغیر طبقه شغلی (jabcat) را به کادر متغیرهای مستقل (Independent list) منتقل کنید.

از کلید Statistics به کادر محاوره آن منتقل شوید.



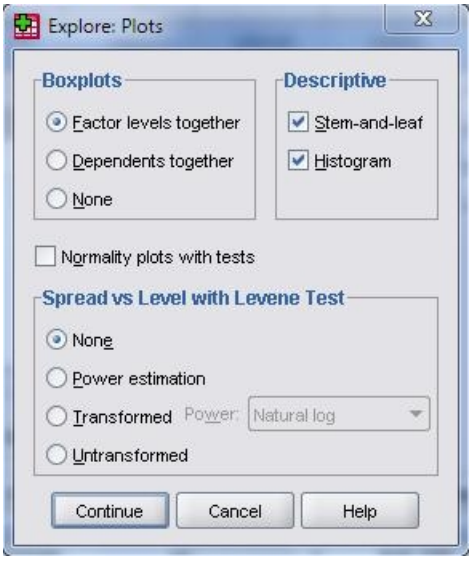
در این کادر محاوره گزینه Descriptive از قبل علامتدار شده تا بعضی از آماره های توصیفی مهم به همراه یک فاصله اطمینان ۹۵٪ برای میانگین حقوق کارمندان در هر گروه شغلی محاسبه شود.

گزینه Outliers را انتخاب کنید تا بیشترین و کمترین حقوق بگیران در هر طبقه شغلی مشخص شوند.

گزینه Precentiles را انتخاب کنید تا چندکهای حقوق کارمندان در هر گروه معلوم شود.

کلید continue را کلیک کنید تا به کادر محاوره اصلی برگردید...

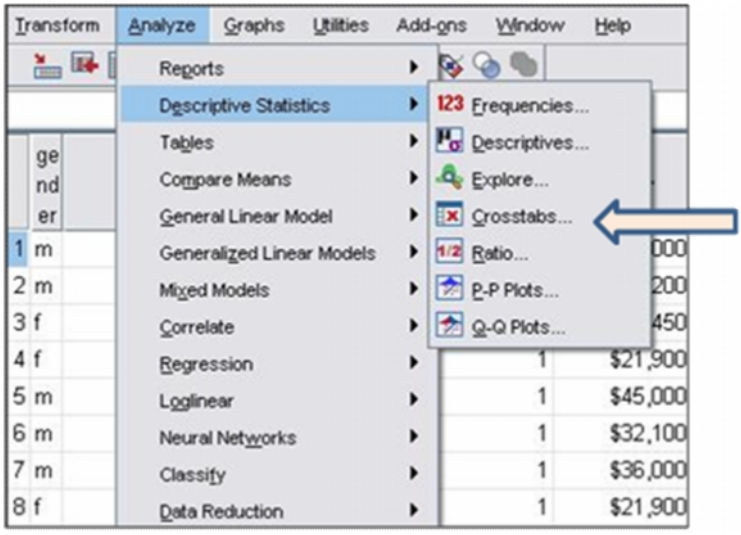
برای رسم نمودار دلخواه، گزینه Plots… را کلیک کنید تا به کادر محاوره Explore: Plots وارد شوید

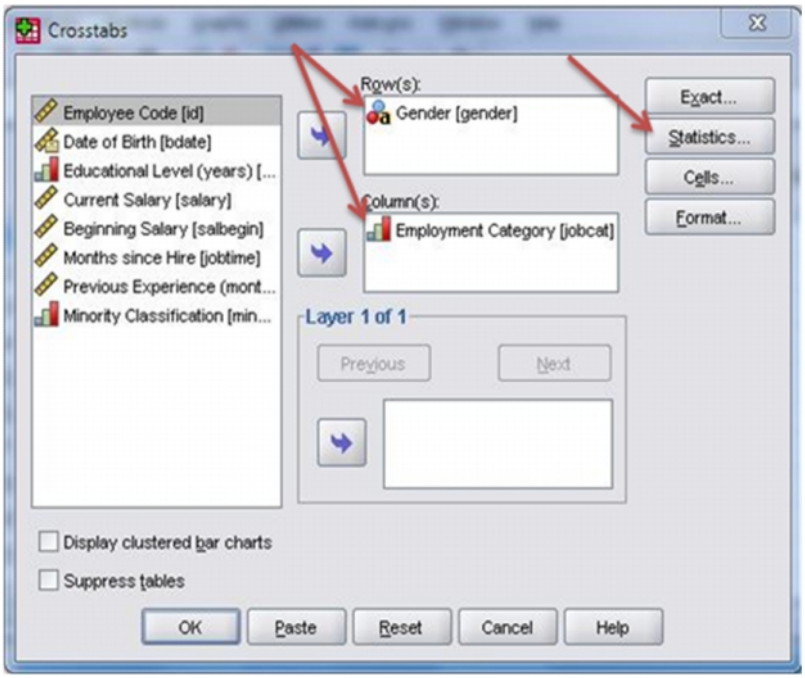


دو گزینه Stem-and-Leaf و Histogram را به منظور رسم نمودار ساقه و برگ و هیستوگرام انتخاب کنید.
کلیدهای continue و ok را کلیک کنید و نتایج را در خروجی مشاهده نمایید.

دستور Crosstabs

جداول توافقی به جداول دو طرفه اطلاق می شود، که در آن داده های آماری در r سطر و c ستون قرار دارند.
در هر جدول توافقی می‌توان تعداد r×c ترکیب مختلف از مقادیر را مشاهده کرد.
در چنین جدولی می‌توان تأثیر یک متغیر کیفی (اسمی یا ترتیبی) را برروی یک متغیر کیفی دیگر بررسی کرد.
در spss، برای ساختن یک جدول فراوانی دوطرفه که به جدول توافقی معروف است، راه‌های زیادی وجود دارد. یکی از این راهها استفاده از دستور Cross Tab است.
می‌توان از منوی اصلی گزینه Analyze و سپس Descriptive Statistics را انتخاب کرده و سپس از دستور Crosstab به کادر محاوره آن وارد شد.

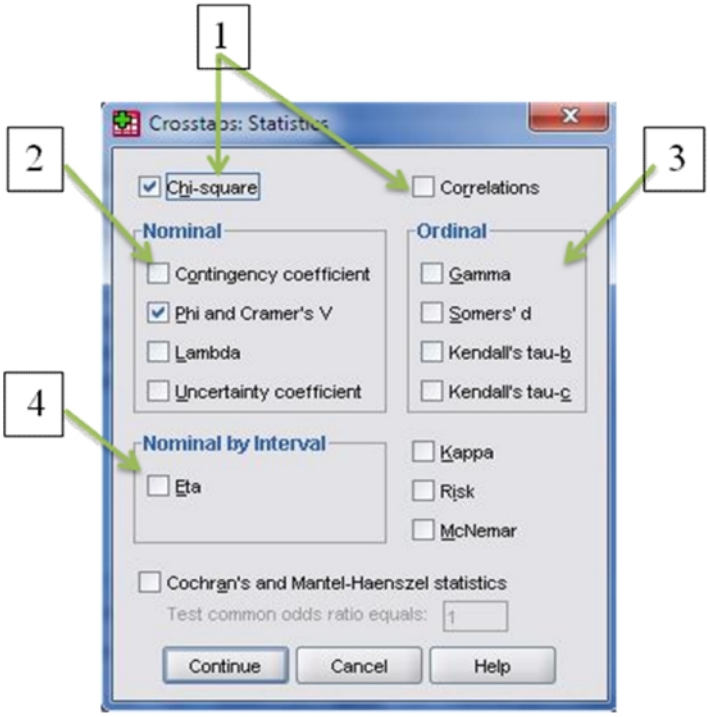




در کادر محاوره Crosstab یک متغیر کیفی به کادر مربوط به سطر Row(s) و یک متغیر کیفی دیگر را به کادر مربوط به ستون Culomn(s) منتقل کنید.

توجه داشته باشید که متغیرهای انتخابی شما باید از نوع اسمی و یا رتبه‌ای باشند و یا اگر متغیر کمی پیوسته هستند آنها را به دسته‌های کوچکتر طبقه بندی کنید.

اگر می‌خواهید ضرایب همبستگی و مقادیر کای‌اسکور را محاسبه کنید، گزینه های مربوط به آنها را در پنجره Statistics انتخاب کنید.



در جدول دوطرفه‌ای که توسط دستور Crosstab بدست می‌آید، می‌توانید ضریب همبستگی (Correlation) یا آماره کای‌اسکور (Chi-square) را محاسبه نمایید.

برای همبستگی متغیرهای اسمی از ضرایبی که در کادر Nominal آمده است استفاده کنید.
برای همبستگی متغیرهای ترتیبی از ضرایبی که در کادر Ordinal آمده است استفاده کنید.
اگر یک متغیر اسمی و دیگری فاصله‌ای است، از ضریب اتا (Eta) استفاده کنید.

تمرین،

داده های 1991 U.S. General Social Survey را از پوشه Sample به آدرس زیر بازکنید.

Program file/ SPSSInc/ SPSS16/ Samples

در این فایل، متغیرهای متفاوتی وجود دارند که تعدادی از آنها را در زیر معرفی می کنیم.

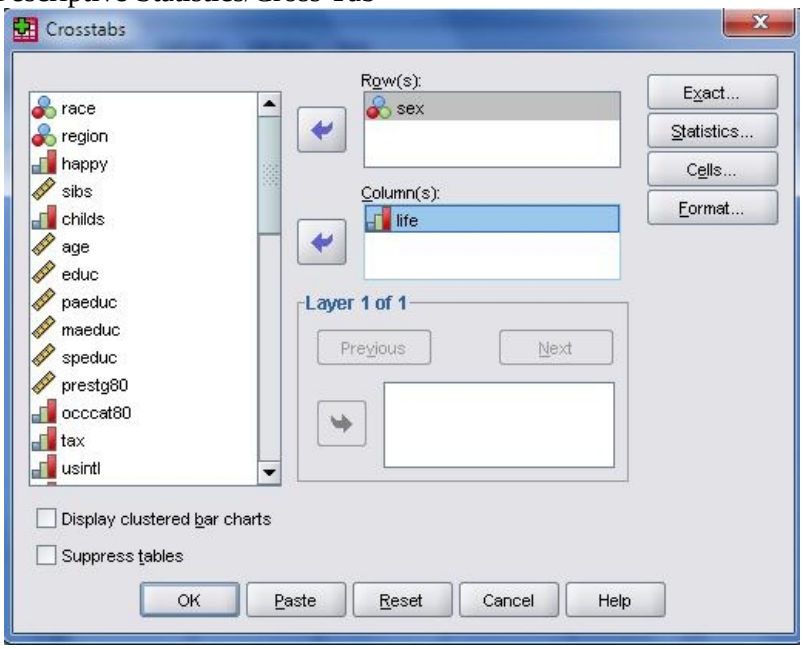
نام متغیر	توضیح	نوع
Sex	جنسیت(مرد- زن)	Nominal
Race	نژاد(سیاه – سفید – سایر نژادها)	Nominal
Region	منطقه محل سکونت (شمال شرقی – شمال غربی- غرب)	Nominal
Life	زندگی را چگونه می بینید (هیجان انگیز – معمولی – خسته کننده)	Ordinal
Sibs	تعداد خواهر و برادر	Scale
Childs	تعداد فرزند	Scale
Age	سن	Scale
Educe	تحصیلات (برحسب سال)	Scale

این داده ها مربوط به یک مطالعه اجتماعی است که بر روی ۱۵۱۷ نفر از جمعیت ۱۸ سال به بالا در سال ۱۹۹۱ در ایالات متحده امریکا انجام شده است.

از این به بعد این فایل داده را با نام مختصر ۱۹۹۱ بکار می‌بریم.

فایل داده های1991 را باز کنید و دستور Crosstab را از فرمان زیر فراخوانی کنید:

Analyze/Descriptive Statistics/Cross Tab



سپس مراحل زیر را دنبال کنید :

در کادر محاوره Crosstab متغیر جنسیت را بـه کـارد سـطرها Row(s) و متغیـر Life را بـه کـادر مربـوط بـه ستون‌ها column(s) ، وارد کنید.

برای رسم نمودار ستونی خوشه‌ای این دو متغیر، گزینه Display Clustered Bar Charts را علامت دار کنید.

گزینه Cells را انتخاب کنید تا به کادر محاوره آن منتقل شوید...

در کادر محاوره باز شده در قسمت Counts گزینه Observed علامتدار شده است، آن را تغییر ندهید.

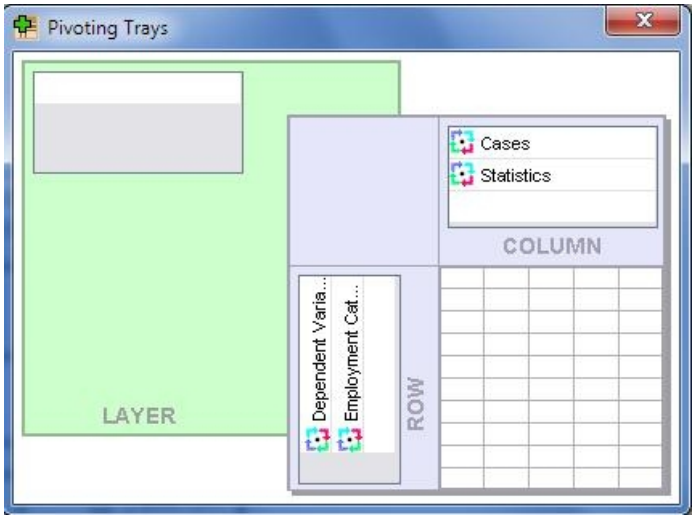
در بخش Percentages گزینه Row را علامت دار کنید.

در پایان کلید Continue و Ok را به ترتیب کلیک کنید و نتایج را در خروجی SPSS مشاهده کنید.

جدول گردان (pivot table)

اگر یک جدول دو طرفه در هر یک از دستورات Explore یا Cross tab برای مقایسه سطوح متغیرها در خروجی SPSS ایجاد کرده‌اید، می‌توانید به سادگی جای سطر ها و ستون ها را عوض کنید و معانی مختلفی از جدول استنباط نمایید. برای این کار روی جدول دو بار کلیک نمایید تا منوی ویراشگر جدول مشاهده گردد.

اینک از منوی Pivot روی گزینه Pivoting Trays کلیک کنید. می توان از همین نوار منو در منوی Veiw گزینه Toolbar را انتخاب نموده تا نوار Formatting Toolbar ظاهر شده سپس بر آیکون جدول گردون کلیک نموده تا کادر Pivoting Trays بازشده جای متغیرهای سطری و ستونی را تغییر می دهیم.



در مکان سطر و ستون این پنجره یک یا چند مربع کوچک به رنگ سبز وقرمز مشاهده می کنید که اگر آنها را جا به جا کنید، جدول خروجی تغییر خواهند کرد. با مشاهده آنها می‌توانید برای سهولت استفاده از اطلاعات جدول، بهترین حالت را به دلخواه انتخاب نمایید.

آمار استنباطی

در آمار استنباطی: با استفاده روشهای آماری خاصی درباره رد یا قبول فرضیه ای تصمیم گیری می نماییم.
تعریف فرضیه: فرضیه اظهار نظری است که در ارتباط با پارامترهای یک یا چند جمعیت بیان میشود.
مثال: فرض کنید محققی ادعا میکندکه میانگین فشار خون بیمارانی که از داروی جدید A استفاده میکنند در مقایسه با بیمارانی که تحت درمان استاندارد B قرار میگیرند کمتر است.
عبارت فوق یک فرضیه است که در ارتباط با میانگین فشار خون(پارامتر مجهول) دو گروه مطرح شده است.

آزمون فرض: با استفاده از آزمون فرضها می توان نسبت به رد یا قبول فرضیه ای تصمیم گیری نمود. به عبارتی در آزمون فرضیه میتوان تعیین نمود که آیا این اضهار نظرها یا فرضیه بیان شده با داده های موجود سازگار است. در این روش بر اساس اطلاعات جمع آوری شده (نمونه گیری) در ارتباط با فرضیه مورد نظر اظهار نظر خواهیم کرد.
در هر آزمون آماری یک فرضیه اولیه وجود دارد که آنرا فرضیه صفر یا H_0 میگویند و بعنوان فرض عدم اختلاف شناخته میشود. در برابر این فرضیه، فرضH_1 یا فرض مقابل وجود دارد (ادعای مطرح شده) که بعنوان فرض وجود تفاوت یا اختلاف شناخته میشود.
در مثال بالا:
فرض صفر یا H_0 : میانگین فشار خون بیمارانی که از درمان A استفاده میکنند با بیمارانی که تحت درمان B قرار میگیرند برابر است.
فرض مقابل یا H_1 : میانگین فشار خون بیمارانی که از درمان A استفاده میکنند با بیمارانی که تحت درمان B قرار میگیرند متفاوت (کمتر) است.

در هر آزمون فرضیه، از دیدگاه آماری، مرتکب دو نوع خطا می شویم:
خطای نوع اول که با α نمایش داده میشود: عبارت است از رد فرض صفر(به اشتباه) وقتی فرض صفر درست باشد. مقدار α در اختیار پژوهشگر است که معمولا 0.05 انتخاب میشود!
خطای نوع دوم که با β نمایش داده میشود: عبارت است از پذیرش فرض صفر(به اشتباه) وقتی فرض مقابل درست باشد.
توان آزمون: عبارت است از رد فرض صفر وقتی فرض مقابل درست باشد (1-β).
در بین دو خطای مطرح شده خطای نوع اول یا α مهمتر از خطای نوع دوم یا β میباشد به همین دلیل ما α را قبل از شروع آزمایش ثابت فرض میکنیم. معمولا مقدار α برابر 0.05 در نظر گرفته میشود.
مفهوم P-value : معمولا نتیجه هر آزمون آماری با P-value بیان میشود. به عنوان مثال اگر متخصص آمار با استفاده از یک روش آماری فرضیه بالا را مورد آزمون قرار داده و P-value را 0.02 بیان کند چه نتیجه ای میگیریم به عبارتی مفهوم 0.02 چیست؟
به ما میگوید که اگر قرار باشد که بر اساس نمونه جمع آوری شده فرض صفر را به اشتباه رد کنیم (درمان A از B بهتر است) احتمال این اشتباه یا خطا تنها ۲ درصد یا ۲ صدم میباشد که این خطا به مراتب کمتر از خطایی است که ما اجازه داریم مرتکب شویم (یعنی $\alpha=0.05$).
و به صورت زیر بیان میشود:

$$P\text{-value}<0.05$$

روشهای آماری متداول در تحلیل اطلاعات

برای آزمون فرضیه در مورد متغیرهایی که دارای دو سطح هستند، سه نوع آزمون t وجود دارد:

t یک نمونه ای، t دو نمونه مستقل، t زوجی

آزمون t یک نمونه ای

از این آزمون زمانی استفاده می‌شود که محقق می‌خواهد بداند،آیا میانگین یک متغیر برابر با مقدار خاصی می باش؟

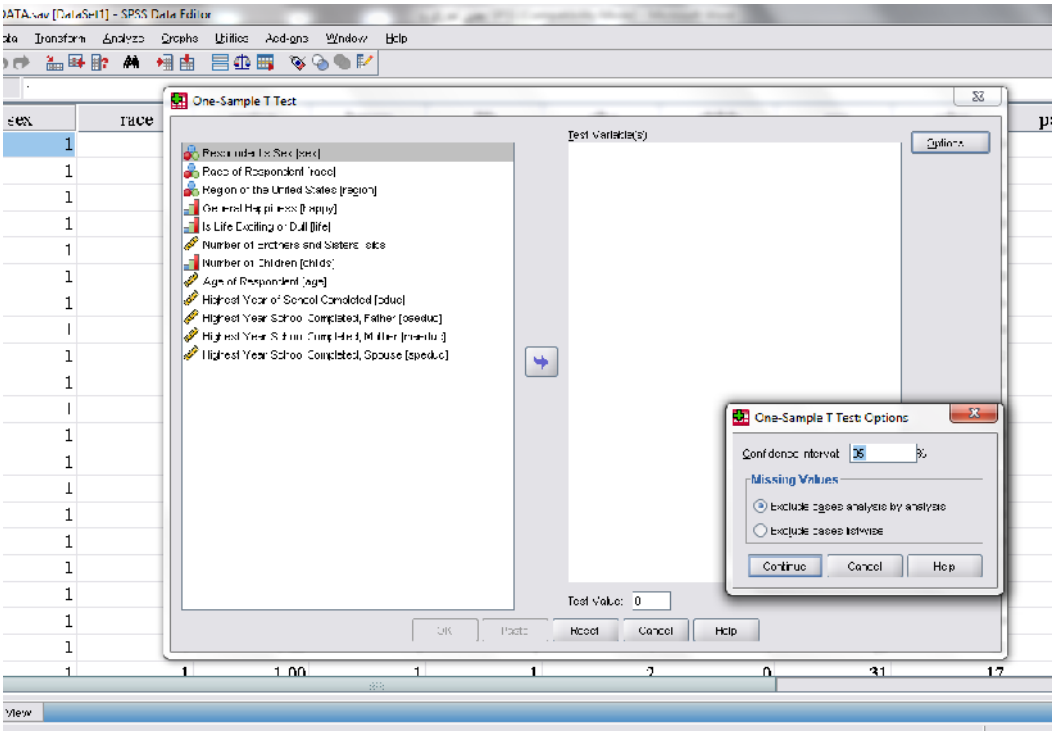
مثال: محقق قصد دارد بررسی نماید آیا یک میانگین نمرات دانشجویان دانشگاه برابر با ۱۴ می باشد یا نه؟

نحوه انجام آزمون t یک نمونه ای

با کلیک بر منوی Analyze ، از پنجره فرعی باز شده بر دکمه Compare Means کلیک کرده، آنگاه بر دکمه One –Sample T-Test کلیک کنید. با انتخاب متغیر مورد آزمون و انتقال آن به کادر مقابل و تعیین عدد مورد نظر برای بررسی فرضیه آزمون انجام می گیرد.

با کلیک بر دکمه Option برآورد فاصله ای در سطح اطمینان تعیین شده توسط شما، محاسبه می گردد.

مثال: یک شرکت تولید کننده صفحه ترمز باید صفحه ترمز هایی تولید کند که قطری معادل ۳۲۲ mm داشته باشند، بخش کنترل کیفیت به صورت تصادفی ۱۶ صفحه از صفحات تولیدی را گرفته و اندازه گیری میکند، برای اطمینان از صحت کیفیت محصولات از آزمون t استفاده میکنیم



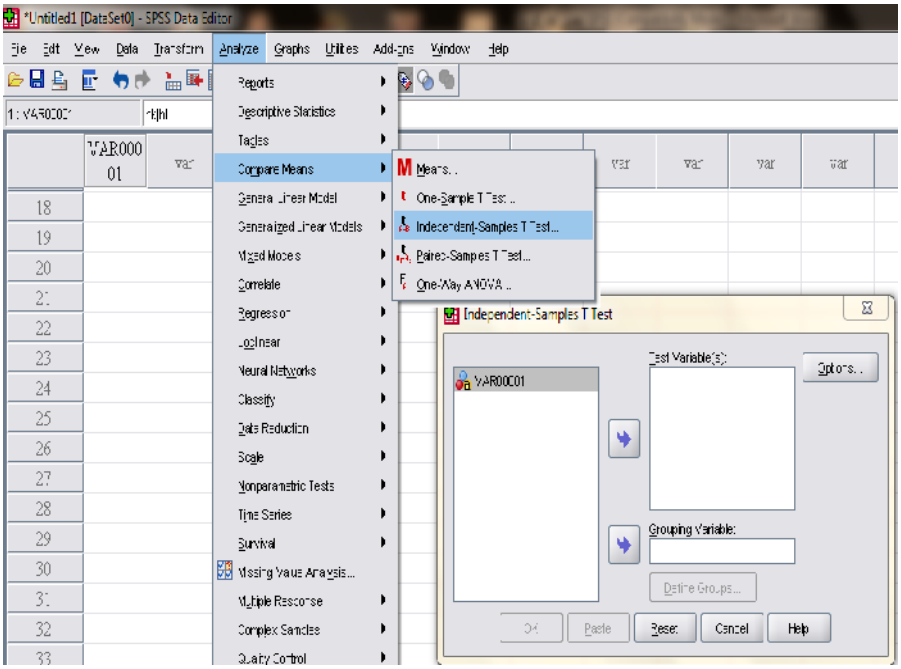
آزمون مقایسه میانگین دو نمونه مستقل t (یا t-test)

این آزمون زمانی که هدف ما مقایسه میانگین دو گروه مستقل میباشد استفاده میشود. مانند مثال بالا.

نحوه انجام آزمون در SPSS

آرایش داده ها در نرم افزار باید اینگونه باشد که، داده های گروه A را در یک ستون وارد نموده و همچنین دادههای گروه دوم یا B را زیر آنها وارد مینماییم سپس متغیر دیگری (ستون دیگری) را در نظر گرفته و در آن ستون در مقابل دادهای ستون اول گروه درمانی A و B را با کدهای ۱و ۲ مشخص مینماییم.

سپس به منوی Analyze رفته و گزینه Compare means را انتخاب نموده و در این قسمت بر روی عبارت 2 Independent Sample کلیک نمایید. سپس متغیر پاسخ یا همان متغیر کمی را در قسمت Test variable و متغیر کد بندی شده را در قسمت Grouping variable وارد نمایید.

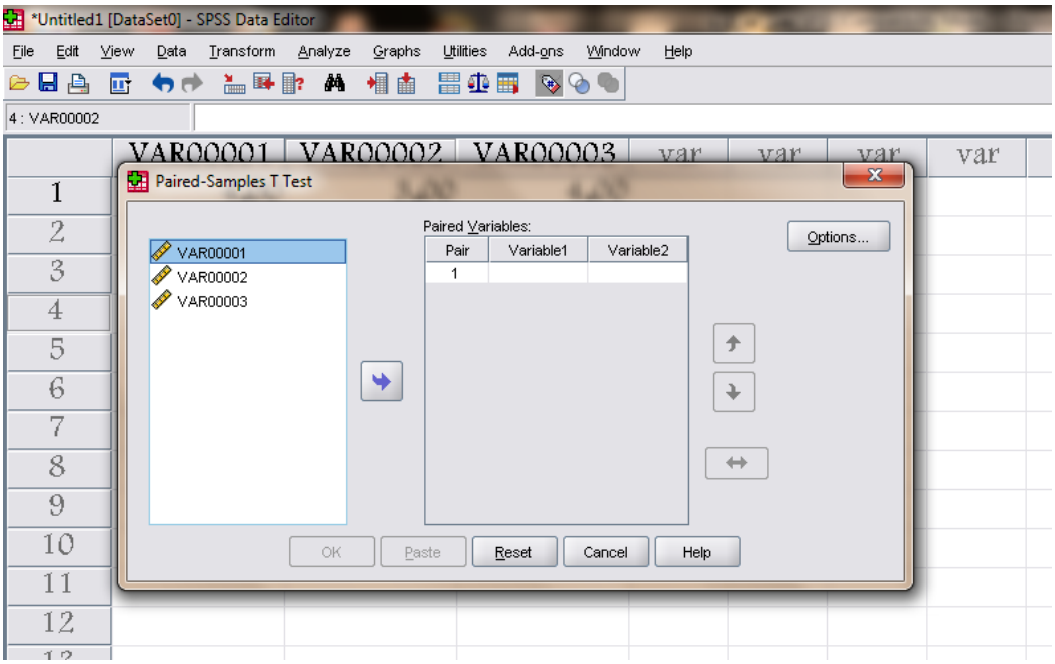


آزمون مقایسه میانگین دو نمونه وابسته(t زوجی Paired t-test)

این آزمون زمانی استفاده میشود که ما یک گروه را دوبار اندازه گیری نماییم. برای مثال فرض کنید که هدف، بررسی اثر یک رژیم غذایی برکاهش وزن یک گروه از افراد باشد. بر اساس آزمون t زوج شده ما ابتدا یک بار وزن افراد را قبل از رژیم غذایی اندازه گیری نموده و سپس یک بار دیگر وزن همان افراد را بعد از رژیم غذایی اندازه گیری مینماییم. در حقیقت در این روش ما هر فرد را کنترل خودش منظور مینماییم. این کار موجب میشود که اگر تفاوتی بین میانگین قبل از رژیم غذایی با میانگین بعد از رژیم غذایی مشاهد شود این تفاوت ناشی از اثر درمان باشد نه عوامل مخدوش کننده دیگر. دلیل این امر نیز آنست که چون هر فرد کنترل خودش محسوب میشود هنگام بررسی اثر درمان، عوامل مخدوش کننده ای مانند سن، جنس، تغذیه، شیوه زندگی و عوامل ژنتیکی برای هر فرد ثابت باقی مانده و در عمل تنها اثر درمان مشاهده میگردد. این در حالیست که در آزمون t-test ما میانگین دو نمونه مستقل را با هم مقایسه مینماییم که در این حالت ممکن است این دو گروه از نظر عواملی که برشمردیم با هم متفاوت بوده و هنگام مقایسه میانگین دو گروه نتایج قابل اعتمادی بدست نیاوریم. بنابر این در شرایطی که این امکان وجود دارد که هر فرد کنترل خودش منظور شود بهتر است ما از روش فوق به منظور طرح ریزی آزمایش خود استفاده نماییم. در هر حال ذکر این نکته ضروری است که جور سازی عملی پر هزینه و وقت گیر بوده و در مواردی عملا انجام آن غیر ممکن است و ما باید بر اساس دو نمونه مستقل آزمایش را طرح ریزی نماییم.

انجام آزمون t زوجی در SPSS

در این روش باید اطلاعات قبل و بعد در دو ستون متفاوت و در مقابل هم همانند آنچه که در جدول بالا مشاهده شد در نرم افزار spss وارد شوند. سپس به منوی Analyze رفته و گزینه Compare means را انتخاب نموده و در این قسمت بر روی عبارت paired sample t-test کلیک نمایید و دو متغیر موجود در سمت راست را به قسمت paired variable در سمت چپ منتقل نموده و دستور OK را اجرا نمایید.



آزمونهای ناپارامتری

نتایج آزمون t زمانی معتبر است که داده ها در هر گروه از توزیع نرمال برخوردار باشند. همچنین اگر حجم نمونه کوچک باشد و نتیجه آزمون t به پذیرش فرض صفر منجر شود (یعنی تفاوت معنی داری بین میانگین دو نمونه مشاهده نشود) ما نمیتوانیم این عدم اختلاف بین دو روش درمانی را به عدم تفاوت واقعی بین درمانها نسبت دهیم. بلکه شاید این مسئله به دلیل حجم نمونه کم در هر گروه باشد.

برای حل این مشکل آزمونهای ناپارامتری مورد استفاده قرار میگیرند. این آزمونها به فرضیاتی مانند حجم نمونه کم، توزیع نرمال و یا فرض برابری واریانسها در هر گروه حساس نمی باشند.

آزمون من ویتنی

این آزمون معادل آزمون t دو نمونه ای مستقل بوده و به جای استفاده از داده های واقعی از رتبه های آنها استفاده میکند.

انجام آزمون من ویتنی باSPSS

روش ورود داده ها دقیقا مانند روش ورود دادها برای انجام آزمون T میباشد. مقادیر صفت مورد بررسی در یک ستون(متغیر) وسپس متغیر دیگری (ستون دیگری) را در نظر گرفته و در آن ستون در مقابل دادهای ستون اول گروه درمانی ۱ و ۲ را با کدهای ۱و ۲ مشخص مینماییم.

سپس به منوی Analyze و گزینه Nonparametric tests را انتخاب مینماییم. سپس وارد منوی 2 Independent Sample شده در پنجره باز شده متغیر مورد نظر در کادر test variableو متغیر گروه بندی را در .Grouping v و سپس در بخش Test type گزینه Mann Whitney را انتخاب مینماییم.

آزمون مک نمار McNemar

آزمون مک نمار برای معنی دار بودن تغییرات به ویژه برای طرحهای قبل و بعد که در آن هر فرد یا آزمودنی به عنوان شاهد خود به کار می رود که در آن متغیر تحت بررسی به صورت اسمی دوحالتی است، استفاده می شود. از این آزمون برای بررسی میزان تاثیر تبلیغات، مطالعه مقاله، کتاب، سخنرانی، دوره آموزشی، ملاقات های فردی و... بر تغییر نظر یا دیدگاه افراد که به دو صورت مثبت و منفی اعلام می شود، استفاده می گردد. چنین آزمایشهایی نمونه های وابسته از داده های اسمی یا ترتیبی را در اختیار می گذارند .

در این گونه آزمون ها فرضیه صفر بیان می دارد که اختلافی بین نظرات آزمون شوندگان در قبل و بعد ایجاد نشده است در مقابل فرضیه یک بین دو حالت اختلاف قائل است .

برای سنجش معنی دار بودن تغییرات با استفاده از این روش باید ابتدا پاسخ های دریافت شده از آزمودنی ها را قبل و بعد در یک جدول دو در دو به صورتی که در زیر مشاهده می کنید قرار داد.

		بعد	
		−	+
قبل	+	**A**	**B**
	−	**C**	**D**

در این جدول آزمودنی هایی که تغییر را نشان میدهند در خانه های A و D هستند. A آنهایی که از + به − و B آنهایی که از − به + تغییر داشته اند و همچنین C و D مربوط به آزمودنی هایی هستند که تغییر نکرده اند. از

آنجایی که A+D مجموع افرادی هستند که در آنها تغییر صورت گرفته است. تحت فرض صفر انتظار این است نیمی از آنان در یک جهت و نیمی دیگر در جهت عکس تغییر کرده باشند

فرض کنید یک روانشناس کودک می خواهد چگونگی آغاز "تماس اجتماعی" در کودکان را بررسی کند. او مشاهده کرده است که کودکانی که تازه وارد مهد کودک می شوند بجای اینکه تماس اجتماعی را با کودکان برقرار کنند با بزرگسالان برقرار می کنند. او پیش بینی می کند که با افزایش میزان آشنایی و تجربه در مهد کودک، کودکان تماس اجتماعی را بیشتر با کودکان برقرار می کنند تا با بزرگسالان.

برای آزمون کردن چنین فرضیه ای او 25 کودک تازه وارد به مهد کودک را مورد مشاهده قرار می دهد و آنها را بر حسب اینکه در اولین روز تماس اجتماعی را با کودکان برقرار می کنند یا بزرگسالان طبقه بندی می کند. پس از گذشت یک ماه مجددا تماس اجتماعی آنها را بررسی و مشاهدات را برای بار دوم ثبت می کند .نتایج مشاهدات قبل و بعد از یک ماه به صورت: 1= تماس با کودکان و2 = تماس با بزرگسالان در جدول زیر آمده است

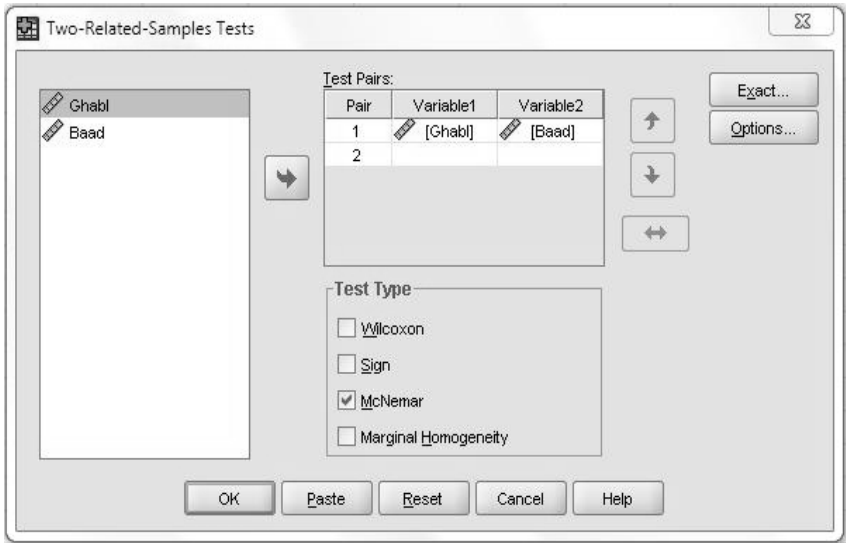
قبل	2	2	1	1	1	2	2	2	1	2	1	2	2	2	1	2	2	2	1	2	2	2	2	2	2
بعد	1	2	1	1	1	2	1	1	2	1	2	2	1	2	1	1	2	2	1	1	1	1	1	1	1

ابتدا داده ها را با تعریف دو متغیر قبل (Ghabl)و بعد (Baad)به SPSS با کدهای تعریف شده برای هر کودک منتقل کنید .

فرمان زیر را اجرا کنید تا کادر محاوره آزمون های دو نمونه ای وابسته باز شود.

Analyze/Nonparametric Tests/2 Related Samples Tests

آزمون مک نمار در spss

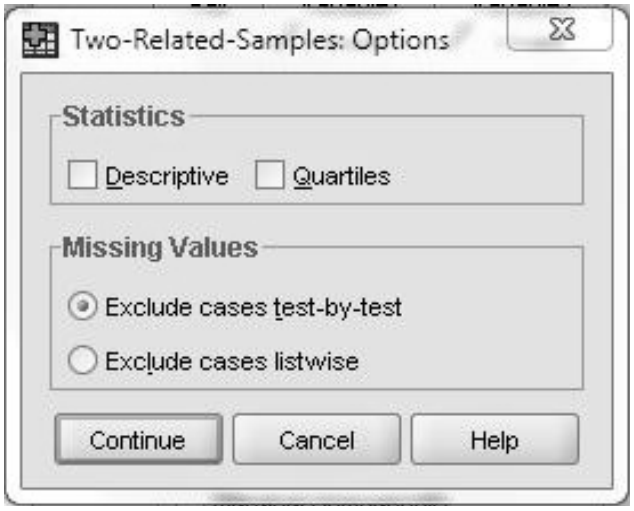


در کادر محاوره به جز آزمون مک نمار می توانید آزمون های ویلکاکسون و آزمون علامت را نیز انجام دهید. البته ما قصد داریم تنها آزمون مک نمار را برای این داده ها انجام دهیم.

در کادر محاوره باز شده، دو متغیر قبل (Ghabl)و بعد (Baad)را متوالیا به کادر .Test Pairsمنتقل کنید. از آنجایی که می شود در کادر محاوره به طور هم زمان بیش از یک آزمون دوتایی انجام داد، شماره های ردیف در کادر Test مربوط به هر یک از آزمون ها است. پس توجه کنید که دو متغیر ورودی باید در یک ردیف قرار گیرند.

برای محاسبه بعضی شاخص های توصیفی، گزینه Optionرا کلیک کنید و در کادر محاوره آن، گزینه Descriptiveرا انتخاب کنید.

در خاتمه OKرا کلیک کنید و نتیجه آزمون را ببینید



نتایج تست McNemar

Ghabl & Baad

Ghabl	Baad	
	1	2
1	3	4
2	14	4

Test Statistics[b]

	Ghabl & Baad
N	25
Exact Sig. (2-tailed)	.031[a]

a. Binomial distribution used.

b. McNemar Test

Sig کمتر از ۰۵/۰ هست فرض صفر رد می شود، مفهوم آن این است که مهد کودک درتماس اجتماعی کودکان تاثیر گذار است. این تاثیر بیشتر در کودکانی است که ابتدا با بزرگسالان تماس برقرارمی کنند ولی بعدا به تماس با کودکان دیگر روی می آورند.

تحلیل واریانس ANOVA

آزمون t تنها در شرایطی که هدف مقایسه میانگین ۲ گروه باشد مورد استفاده قرار میگیرد، روشی که برای مقایسه میانگین بیش از دو گروه به کار می رود یعنی وقتی بخواهیم میانگین های یک صفت کمی را درسه یا بیش از سه گروه مقایسه کنیم، تجزیه و تحلیل واریانس(***ANOVA***) نامیده می شود. در واقع این تجزیه و تحلیل، ما را در فهم تفاوت بین گروه ها یاری می کند. از مزایای استفاده از این آزمون این است که تنها با انجام یک بار آزمودن، اختلاف میان میانگینهای کلیه تیمارهای موجود در آزمایش، مورد بررسی قرار می گیرد.

آزمون فرضیه برای این هدف بصورت زیر می باشد:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_k \\ H_1: \mu_i \neq \mu_j \quad \text{حداقل یکی از } \mu_i \text{ با سایر آنها تفاوت داشته باشد} \end{cases}$$

برای انجام آنالیز واریانس نیز همانند سایر آزمون های پارامتری باید پیش فرض هایی برقرار باشد؛ این پیش فرض ها به قرار زیرند:

نمونه های گرفته شده از هر جامعه کاملا تصادفی و مستقل باشند.

متغیر تصادفی مورد نظر (صفت مورد بررسی) در هر کدام از جامعه ها از توزیع نرمال پیروی کنند.

واریانس متغیر تصادفی در همه ی جامعه ها برابر باشند.

نمونه های مختلف از جوامعی گرفته شده اند که تنها از یک روش طبقه بندی شده باشند.

روش آنالیز واریانس ANOVA بر مبنای آزمون F انجام خواهد شد. آزمون مقایسه میانگین سه گروه یا بیشتر بر مبنای جدول زیر که جدول تحلیل واریانس نام دارد میباشد.

منبع	مجموع توانهای دوم	درجه آزادی	MS(Mean square)	F
درمانها یا بین گروها	$SSB = \sum_{j=1}^{k} n_j (\bar{y}_i - \bar{y})^2$	K-1	$MSB = \frac{SST}{k-1}$	$F = \frac{MSB}{MSW}$
درون گروهها	$SSW = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_i)^2$	N-K	$MSW = \frac{SSW}{N-K}$	
کل	$SST = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$	N-1		

اگر در این جدول مقدار F مشاهده شده بزرگتر از مقدار بحرانی جدولF باشد فرض صفر رد خواهد شد. یعنی اگر $F_{obs} > F_{\alpha}(k-1, N-k)$ باشد آنگاه فرض صفر رد میشود. البته در نرم افزار SPSS علاوه بر مقدار آماره F احتمال معنی داری یا همان P-value نیز گزارش میگردد. درچنین شرایطی P-value کوچکتر از α (0.05 یا0.01) معنی دار در نظر گرفته خواهد شد. در هر حال در هنگام استفاده از جدول ANOVA باید به این نکته مهم توجه داشته باشیم که اگر $F_{obs} > F_{\alpha}(k-1, N-k)$ یا $P-value < 0.05$ باشد ما تنها نتیجه میگیریم که حداقل دو گروه از گروههای مورد مطالعه از نظر آماری تفاوت معنی داری را نشان میدهند اما نمیتوانیم تشخیص دهیم که این گروهایی که از نظر آماری متفاوتند کدام گروهها خواهند بود. در چنین شرایطی ما نیاز داریم که از آزمونهای پسین به منظور شناسایی این تفاوتها استفاده نماییم.

روش آزمون آنالیز واریانس را با انجام یک مثال بیان می کنیم.

مثال: متوسط زمان بستری شدن بر ای نمونه های ده تایی از بیماران در ماه شهریور برای یک بیماری خاص در ۵ بیمارستان به صورت زیر می باشد؛ بررسی کنید که آیا میان متوسط زمان بستری شدن(به روز) بیماران ۵ بیمارستان تفاوت معنی داری وجود دارد یا نه؟ در صورت وجود اختلاف نشان دهید که میان کدام بیمارستانها در این زمینه تفاوت وجود دارد.

جدول ۱

بیمارستان A	7	7	8	6	7	5	8	7	6	8	μ_1
بیمارستان B	8	8	8	8	7	7	6	6	6	5	μ_2
بیمارستان C	7	5	5	5	4	7	4	4	5	5	μ_3
بیمارستانD	8	9	9	11	6	10	11	11	10	12	μ_4
بیمارستانE	4	9	6	4	4	4	5	5	4	6	μ_5

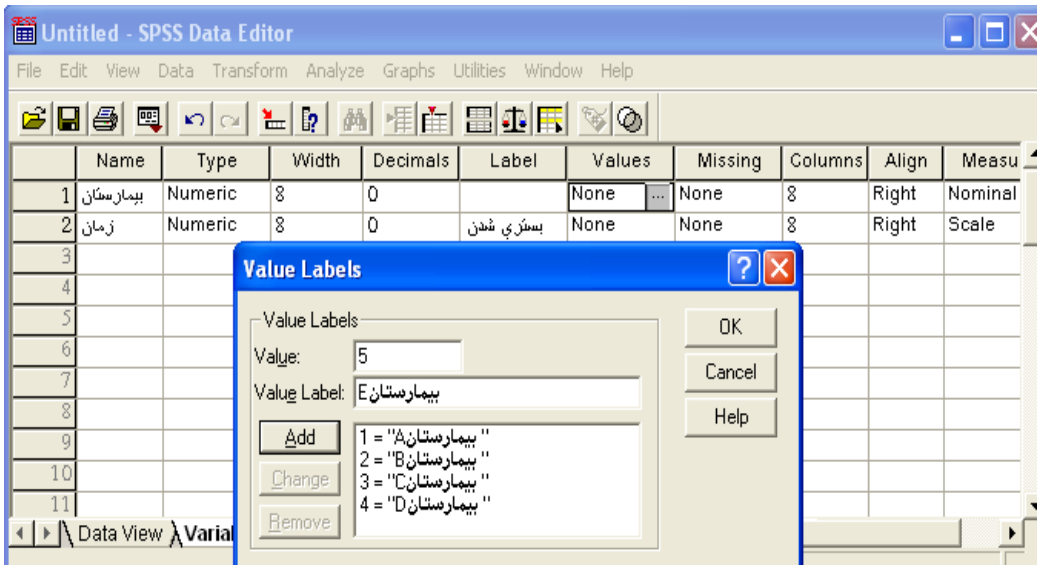
تیمار= متوسط زمان بستری شدن بیماران

روش ورود داده ها برای انجام آزمون ANOVA در نرم افزار SPSS

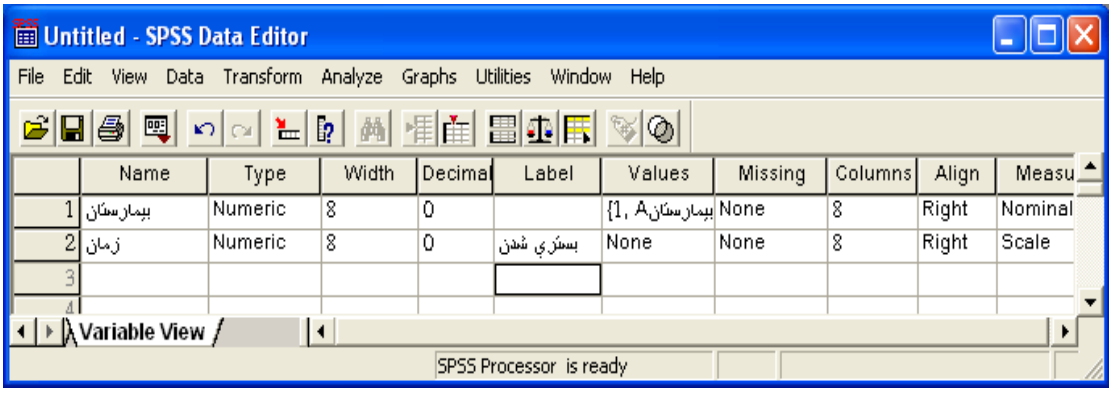
روش ورود دادها برای انجام این آزمون دقیقا مانند روش t-test است. یعنی ابتدا دادهای گروه ۱ را در ستون اول وارد نموده و سپس دادهای گروه دوم و سوم (اگر سه گروه را مقایسه می کنید) را زیر آنها وارد نمایید. سپس متغیر دیگری (ستون دیگری) را در نظر گرفته و در آن ستون، در مقابل دادهای ستون اول متغیر نام بیمارستان را با کدهای ۱و ۲ و ۳ و ... مشخص نمایید. بر این اساس در نرم افزار ابتدا در پنجره ***Variable View*** دو متغیر به نام نام بیمارستان و دیگری زمان بستری شدن بیمار تعریف می کنیم و سپس مانند توضیحاتی که در آغاز گفته شد ستونهای مورد نظر را متناسب با نوع متغیر تنظیم می کنیم.

در این قسمت برای متغیر نوع بیمارستان در ستون ***Values*** نوع بیمارستانها را تعریف می کنیم.

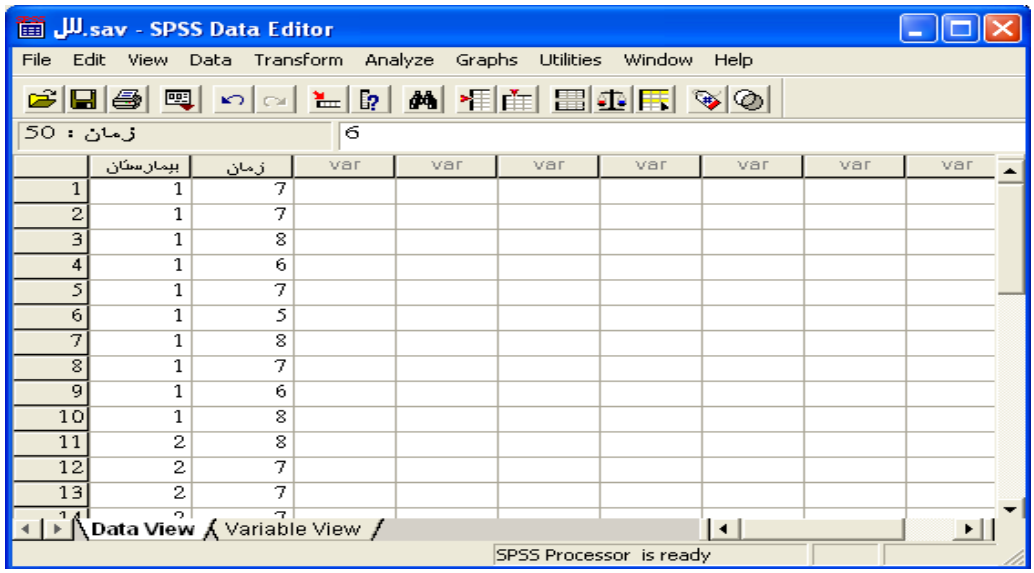
(در پنجره ***Data View*** به جای اسامی بیمارستان از کدهای (۱ و ۲ و.....) استفاده می کنیم).



و در پایان با کلیک کردن بر روی ***ok*** پنجره ***Variable View*** به صورت زیر در می آید.

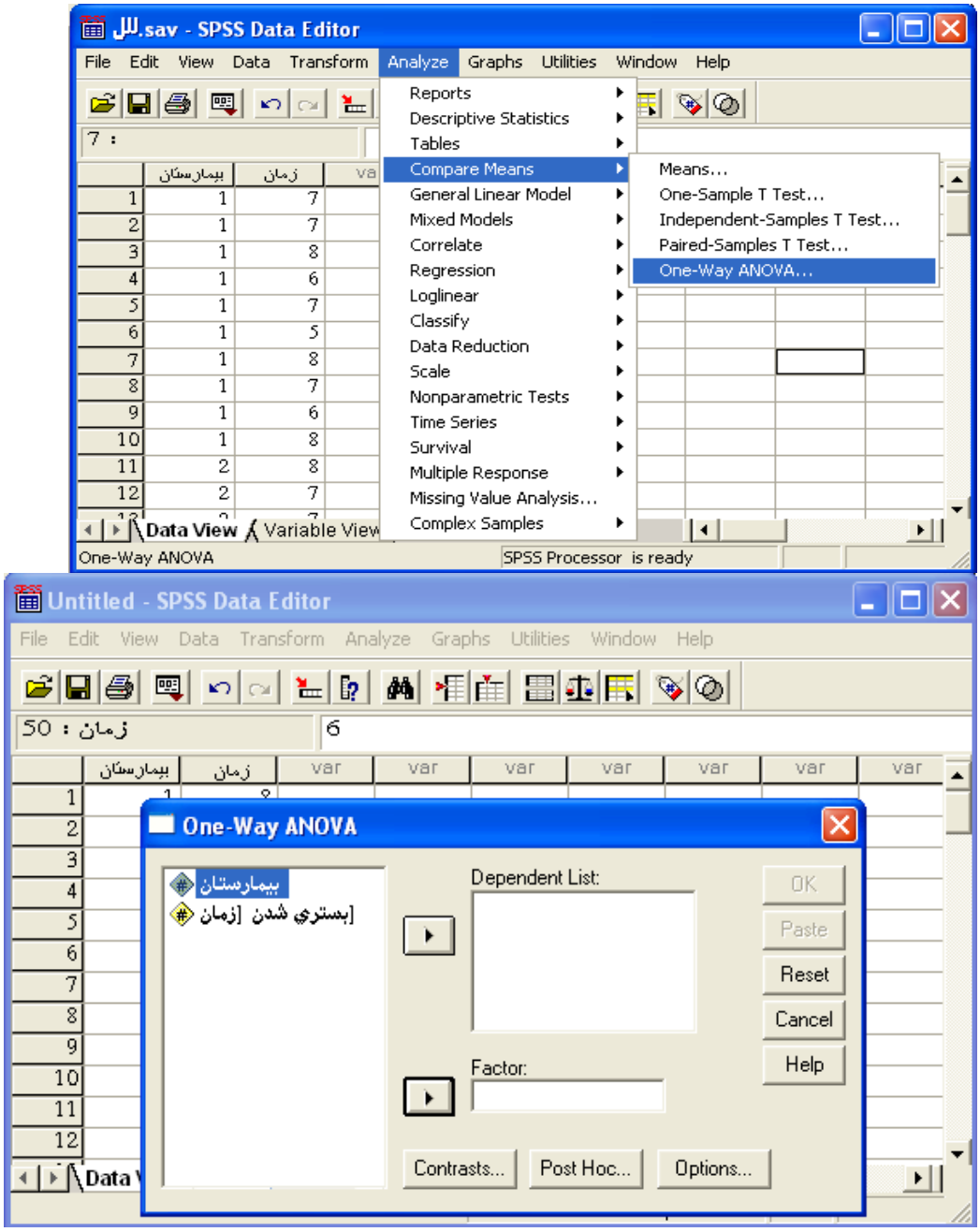


در پنجره ***Data View*** داده ها را به صورت زیر وارد می کنیم. در ستون بیمارستان کد مربوط به بیمارستانها را (۱ ۵) وارد کرده و جلوی هر کد در ستون زمان، مدت زمان بستری شدن بیماران بیمارستانهای مختلف را طبق جدول ۱ وارد می کنیم.

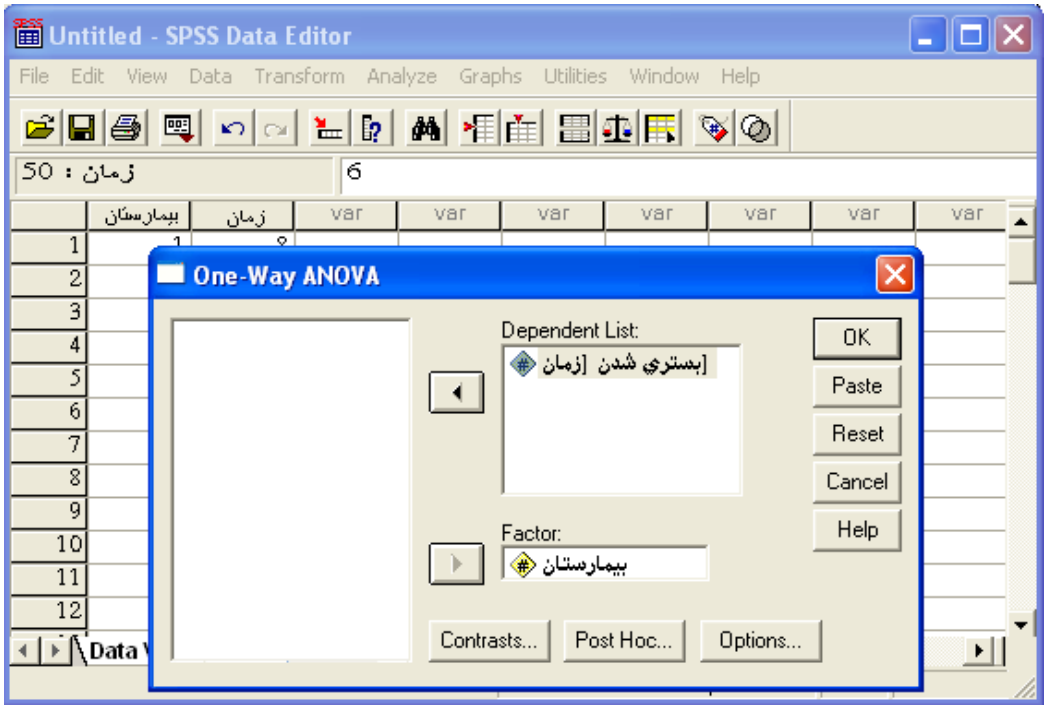


برای انجام آزمون به صورت زیر عمل می کنیم:

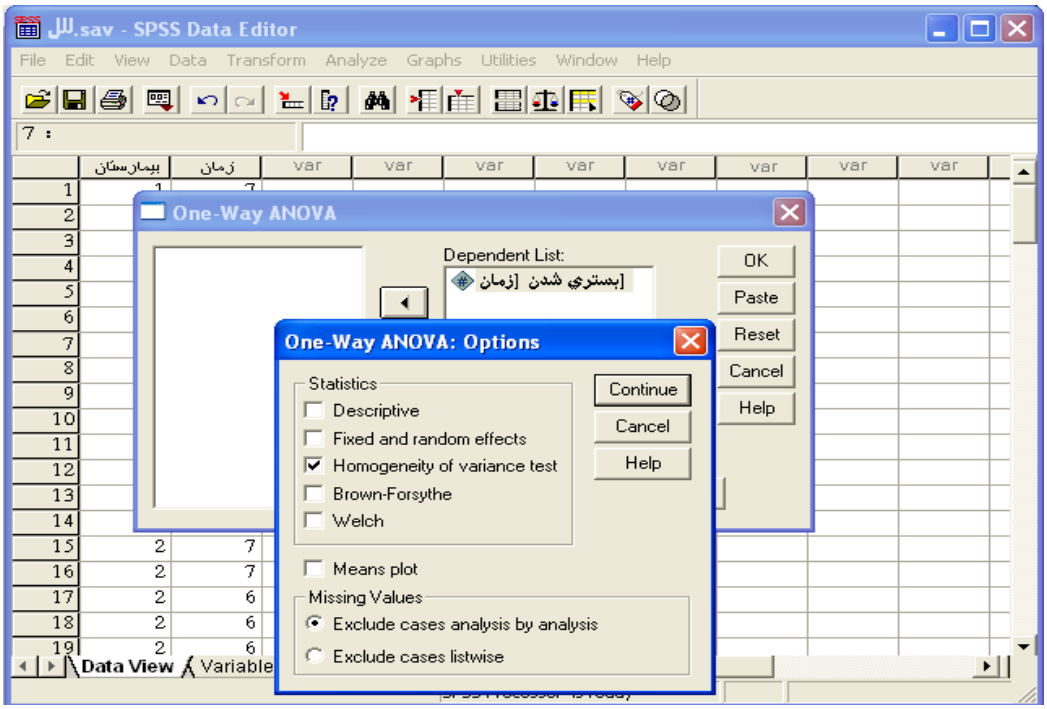
ابتدا از نوار منو SPSS به منوی Analyze رفته و گزینه Compare means را انتخاب نموده و در این قسمت بر روی عبارت one-way ANOVA کلیک نمایید. سپس در کادر Dependent List متغیر کمی مورد نظر و در قسمت Factor متغیر گروه بندی شده (کد بندی) را وارد نمایید.



سپس با کلیک کردن بر روی دو متغیر و با استفاده از مثلث های سیاه کوچک آنها را به مستطیلهای سمت راست منتقل می کنیم.



قبل از اینکه بر روی کلمه ***ok*** کلیک کنیم برای بررسی اینکه آیا بین واریانسهای (مدت زمان بستری شدن بیمـاران) ۵ بیمارستان تفاوت وجود دارد یا خیر بر روی دکمه ***option*** کلیک می کنیم تا پنجره زیر باز شود.



در پنجره ***One-Way ANOVA:Option*** گزینـه ***Homogeneity of variance test*** را فعـال کـرده، در ادامـه ابتدا بر روی کلمه ***Continue*** و سپس ***Ok*** کلیک می کنیم تا خروجیهای زیر بدست آیند.

آزمون همگنی واریانسها (جدول ۲)

آزمون برابری میانگینها (جدول ۳)

نتایج بدست آمده از جدول ۲ نشان می دهد که در آزمون مقایسه بین واریانسهای ***مدت زمان بستری شدن*** ۵ بیمارستان اختلاف معنی داری از نظر آماری وجود ندارد (P-value>0.05). اما نتایج بدست آمده از جدول ۳ نشان می دهد که میان میانگینهای ***مدت زمان بستری شدن*** ۵ بیمارستان اختلاف معنی داری وجود دارد. (P-value <0.05).

نکته: در تجزیه و تحلیل آنالیز واریانس، اگر آزمون معنی دار نگردد و فرض H_o پذیرفته شود، تجزیه و تحلیل به پایان می رسد و نشاندهنده این موضوع می باشد که میان تیمارهای (میانگینها) گروههای تفاوتی وجود ندارد.
اما اگر فرض H_o رد شود نشاندهنده اختلاف میان تیمارها می باشد و ما تنها نتیجه میگیریم که حداقل دو گروه از گروههای مورد مطالعه از نظر آماری تفاوت معنی داری را نشان میدهند اما نمیتوانیم تشخیص دهیم که این گروهایی که از نظر آماری متفاوتند کدام گروهها خواهند بود. در چنین شرایطی ما نیاز داریم که از آزمونهای تعقیبی به منظور شناسایی این تفاوتها استفاده نماییم.

آزمونهای تعقیبی:

به دلیل اینکه داده ها نشان می دهند که میانگینهای ۵ بیمارستان با هم تفاوت معنی داری دارند در نتیجه به دنبال تشخیص اختلاف بین گروهها (بیمارستانها) هستیم تا ببینیم کدام دو بیمارستان (گروه) اختلاف دارند.
بدین منظور مسیر بالا را دوباره تکرار می کنیم و به جای کلیک بر روی گزینه ***Option*** گزینه ***Post Hoc*** را انتخاب می کنیم تا پنجره ***One- Way ANOVA: Post Hoc Multiple Comparisons*** باز شود.
در این پنجره انواع آزمونهایی را که می توانیم برای مقایسه میانگینها مورد استفاده قرار دهیم آورده شده است.
این پنجره از دو بخش تقسیم شده است.

قسمت بالا مربوط به آزمونهای مورد استفاده در حالتی که واریانس جوامع تفاوتی نداشته باشند:

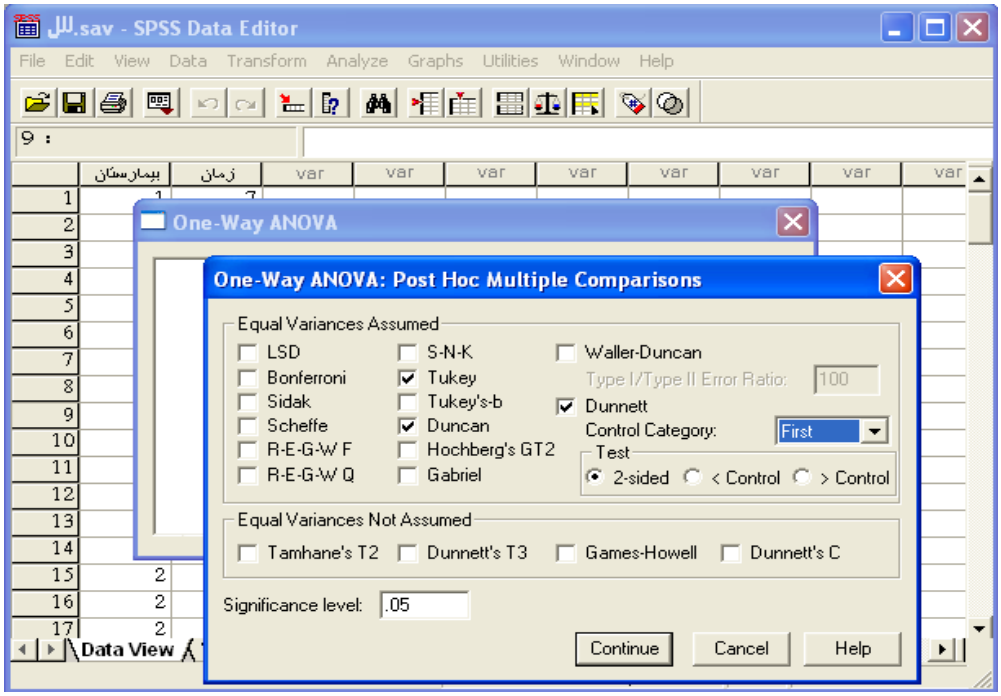
(Equal Variances Assumed)

قسمت پایین مربوط به آزمونهای مورد استفاده در حالتی که واریانس جوامع متفاوت باشند:

(Equal Variances Not Assumed)

در این مثال چون فرض همگنی واریانسها پذیرفته شد، به بیان دیگر فرض H_0 کـه برابـری واریانسـها را مطـرح مـی کند رد نشده شده است از آزمونهای بالایی استفاده می کنیم.

در این قسمت چند مورد از مهمترین آزمونها را مورد بررسی قرار می دهیم.



در پنجره حاضر در حالتی که واریانس گروهها با هم اختلاف معنی داری ندارند به طور معمول رایج ترین روش های مقایسه های چند گانه عبارتند از:

- روش کمترین مربعات LSD
- روش بونفرونی Bonferroni
- روش توکی Tukey
- روش شفه Scheffe
- روش نیومن کلز Newman-Keuls
- روش دانت Dunnett's Procedure

که به کار گیری این روش ها، غالبا به نتایج مشابهی ختم می شود. اما باید توجه کرد که کاربرد روش دانت در مواقعی است که یکی از گروه ها، گروه شاهد یا (Control) باشد.

در اینجا سه آزمون متداول (***Dunnett*** ،***Duncan***،***Tukey***) را برای تجزیه و تحلیل آماری انتخاب می کنیم.

در مورد آزمون ***Dunnett*** این نکته را باید مورد توجه قرار داد که از میان یکی از گروهها (۵ بیمارستان) یکی را به عنوان گروه کنترل (شاهد) در نظر می گیریم و سایر گروهها را با آن می سنجیم. این گروه می توان گروه اول(***First***) یا گروه آخر (***Last***) باشد. انتخاب هر کدام از این گروهها به عنوان گروه اول یا آخر در نتایج تغییری ایجاد نمی کند. برای این کار با فعال کردن آزمون ***Dunnett*** گزینه ***Control Category*** فعال شده و در مربع روبروی آن گروه کنترل را انتخاب می کنیم. در مرحله بعد با کلیک بر روی ***Continue*** به پنجره قبلی می رویم و با کلیک بر روی ***Ok*** می توانیم خروجیهای مورد نظر را مشاهده کنیم.

Post Hoc Tests

Homogeneous Subsets

برای نتیجه گرفتن درباره وضعیت برابری یا عدم برابری میانگینها از دو ستون 95% Confidence Interval و Sig استفاده می کنیم. عدد مشاهده شده در ستون Sig معرف P-Valuse بدست آمده در آزمونها می باشد و چون آزمونها در سطح ۰/۰۵ مورد بررسی قرار می گیرند مبنای پذیرش یا عدم پذیرش آنها یعنی قبول یا رد فرض اولیه H_o مقایسه با مقدار ۰/۰۵ می باشد.

در آزمونها اگر Sig<0.05 فرض H_o رد می شود و اگر Sig>0.05 فرض H_o رد نمی شود.

هم ارز با ستون Sig ستون مربوط به فاصله اطمینان(95% Confidence Interval) است که نمایش دهنده یک فاصله اطمینان ۹۵ درصدی و همچنین تأییدی بر نتایج بدست آمده در ستون Sig می باشد.

اگر p- values بدست آمده را با α نشان دهیم رابطه زیر برقرار است: ضریب اطمینان $= 1-\alpha$
بنابراین وقتی مقدار α یعنی سطح معنی داری برابر ۰/۰۵ باشد ضریب اطمینان ۰/۹۵ می شود.

اگر فاصله اطمینان که یک حد پایین(L) و یک حد بالا(U) دارد، عدد صفر را شامل شود(L<0<U) نشاندهنده این است که فرض H_o یعنی برابری میانگینها رد نمی شود و این هم ارز Sig>0.05 می باشد و اگر این بازه شامل صفر نباشد هم ارز این است که Sig<0.05 و بیان می کند که فرض H_o یعنی برابری میانگینها رد می شود.

در خروجی مربوط به آزمون Dunnett مشاهده می کنیم: بیمارستان A که به عنوان گروه کنترل در نظر گرفته شد با تک تک بیمارستانها از نظر میانگین مدت زمان بستری شدن مقایسه آماری شد و نتایج در جدول آورده شده است. در ستون مربوط به Sig که مقادیر P-Valuse را برای هر ازمون جداگانه نشان می دهد، هر جا Sig<0.05 باشد نشان می دهد که فرض برابری دو میانگین رد شده است. به بیان دیگر بین میانگین مدت زمان بستری شدن برای دو بیمارستان تفاوت معنی داری وجود دارد.

درخروجی مربوط به آزمون Tukey تک تک بیمارستانها با هم مقایسه می شوند. در این آازمون به علت یک دامنه بودن هر کجا Sig<0.05 فرض برابری میانگینها رد می شود(به بیان دیگر بین میانگین مدت زمان بستری شدن دو بیمارستان از نظر آماری تفاوت معنی داری وجود دارد.)

نتایج نشان می دهند که میانگین بیمارستان A با بیمارستان B تفاوت ندارد(Sig>0.05). اما میان میانگین بیمارستان A با میانگین سایر بیمارستانها تفاوت معنی داری وجود دارد(Sig<0.05).

در خروجی مربوط به زیرمجموعه های همگن (Homogeneous Subsets) آزمونهای توکی و دانکن میانگینهایی که با هم تفاوت ندارند را در یک زیر گروه قرار می دهند.

به طور معمول نتایج بدست آمده از آزمون ها هر کدام تأییدی بر نتیجه آزمون دیگر است. البته با توجه به درجه دقت آزمونها در بعضی مواقع ممکن است که نتیجه بدست آمده در یک آزمون با نتیجه بدست آمده در آزمون دیگر متفاوت باشد.

نکته: نتایج آزمونANOVA زمانی معتبر است که داده ها در هر گروه از توزیع **نرمال** برخوردار باشند. البته ذکر این نکته ضروری است که فرض توزیع نرمال معمولا در شرایطی که حجم نمونه کوچک است برقرار نمیباشد. بنابر این در چنین شراطی استفاده از آزمون ANOVA به منظور مقایسه میانگین نمونه سه گروه (یا بیشتر) مستقل چندان قابل اعتمادنخواهد بود. در چنین شرایطی به جای استفاده از آزمون ANOVA از آزمون ناپارامتری معادل آن یعنی Kruskal-Wallis H test استفاده مینماییم.

نرمالیتی داده ها

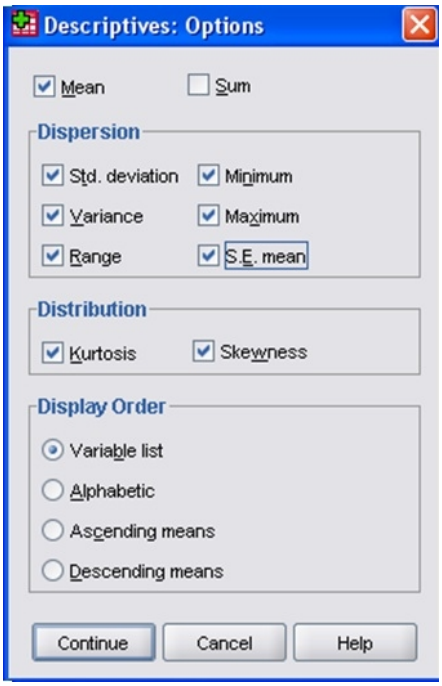
برای چک کردن نرمالیتی داده ها چندین گام وجود دارد:

گام اول:

ابتدا چولگی و کشیدگی داده هایتان را آزمون کنید. این کار را می توان از مسیر زیر در SPSS انجام داد:

Analyze> Descriptive Statistics> Descriptives

در کادر محاوره ای که باز می شود متغیر هایی که می خواهید چولگی و کشیدگی آن را آزمون کنید را به کادر سفید انتقال دهید. سپس روی کلید options کلیک کنید و در کادر محاوره ی آن گزینه های Skewness و kurtosis را فعال کنید .



چولگی برابر با گشتاور سوم نرمال شده است. چولگی در حقیقت معیاری از وجود یا عدم تقارن تابع توزیع می باشد. برای یک توزیع کاملاً متقارن چولگی صفر و برای یک توزیع نامتقارن با کشیدگی به سمت مقادیر بالاتر چولگی مثبت و برای توزیع نامتقارن با کشیدگی به سمت مقادیر کوچکتر مقدار چولگی منفی است. در شکل زیر چولگی مثبت و منفی را می بینید.

کشیدگی یا کورتزیس نشان دهنده قله‌مندی یک توزیع است. مقدار کشیدگی را با گشتاور چهارم نرمال بر آورد کرده اند، به عبارت دیگر کشیدگی معیاری از تیزی منحنی در نقطه ماکزیمم است و مقدار کشیدگی برای توزیع نرمال برابر ۳ می باشد. کشیدگی مثبت یعنی قله ی توزیع مورد نظر از توزیع نرمال بالاتر و کشیدگی منفی نشانه ی پایین تر بودن قله از توزیع نرمال است.

در حالت کلی معمولا چنان چه چولگی و کشیدگی در بازه ی (۲ ، ۲) نباشند داده ها از توزیع نرمال بسیار دور بوده و می بایست قبل از هر گونه آزمونی که برای انجامشان باید فرض نرمال بودن داده ها برقرار باشند؛ اصلاح گردند.(البته ممکن است بعضی از آمار دادنان این بازه را کوچکتر یا بزرگتر در نظر بگیرند.)

مثال زیر را در نظر بگیرید.

از سری داده های آماده ی SPSS فایل adl.sav را از مسیر زیر باز کنید:

Open> data>

در کادر open data در look in به آدرس زیر رفته و فایل adl.sav را انتخاب کنید:

C:\ > program files> SPSSInc > SPSS> Samples

می خواهیم نرمال بودن داده های سن (age)و Hopital LOS(los) را چک کنیم: برای این کار از مسیر * کادر Descriptiveرا باز کنید و این دو متغیر را به کادر سفید انتقال دهید و در منوی optins دو گزینه Kurtosis و Skewnessرا فعال کنید .و در نهایت دکمه ی ok را بزنید. با این کار خروجی زیر را دریافت می کنید:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pt. age	100	66	91	71.76	3.944	1.483	.241	4.779	.478
Hospital LOS	100	12	25	17.25	2.595	.094	.241	-.192	.478
Valid N (listwise)	100								

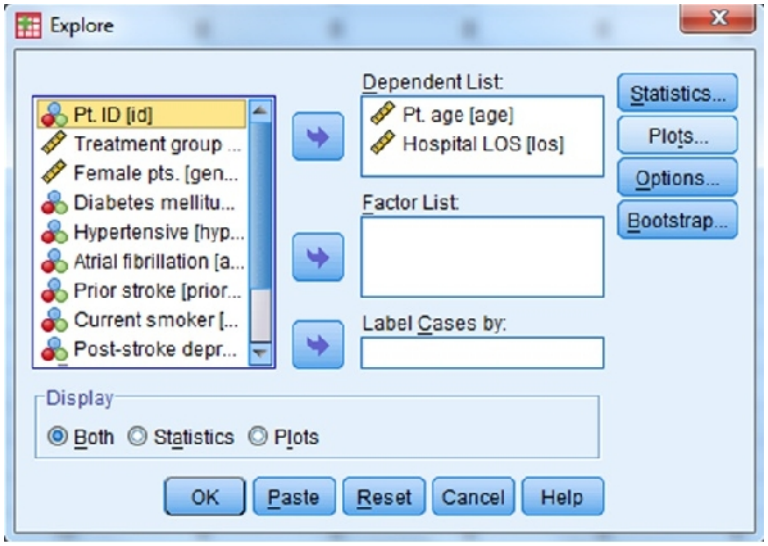
مقدار چولگی مشاهده شده برای متغیر سن ۱.۴۸۳ است این می تواند ما را به نرمال بودن توزیع این متغیر امیدوار کند یعنی از لحاظ کجی متغیر سن مانند نرمال بوده و توزیع آن متقارن است اما مقدار کشیدگی آن از ۳ بیشتر است و این می رساند که قله ی این توزیع از نرمال بالا تر قرار می گیرد. لذا این متغیر با اینکه دارای توزیع متقارنی است اما نرمال نخواهد بود.

اما مقدار چولگی و کشیدگی برای متغیر los در بازه ی (-2 , 2)قرار داشته و می توان گفت که این متغیر می تواند نرمال باشد.

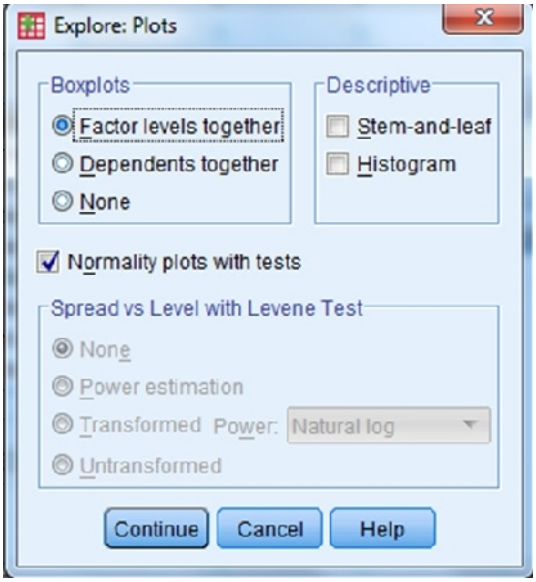
گام دوم:

پس از بررسی عادی یا نرمال بودن کشیدگی و یا چولگی توزیع داده هایتان، به سراغ آزمون شاپیرو ویلک بروید تا از نرمال بودن داده هایتان مطمئن گردید. برای این کار از مسیر زیر گزینه Explore را انتخاب تا وارد کادر محاوره ی زیر شود.

Explore Descriptive Statistics > > Analyze



در مثال بالا همان دو متغیر سن age و los را مطابق شکل وارد لیست متغیر های وابسته کنید و سایر جاها را خالی بگذارید. سپس به منوی plots رفته و گزینه ی Normality plots with tests را تیک دار کنید.



با این عمل خروجی شما شامل جدولی است تحت عنوان Tests of Normality که به شما دو مقدار سطح معناداری را برای هر کدام از متغیر ها به طور مجزا می دهد. این مقادیر در تشخیص نرمالیتی داده ها بسیار تعیین کننده است.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pt. age	.125	100	.001	.891	100	.000
Hospital LOS	.122	100	.001	.975	100	.054

a. Lilliefors Significance Correction

معمولا چنانچه سطح معناداری در آزمون Shapiro-Wilk که در این جدول با sig. نمایش داده می شود بیشتر از ۰.۰۵ باشد می توان داده ها را با اطمینان بالایی نرمال فرض کرد. در غیر این صورت نمی توان گفت که داده ها توزیعشان نرمال است. با توجه به جدول فوق و مقادیر سطح معناداری برای متغیرهای age وlos می توان گفت که توزیع متغیر los می تواند با احتمال خوبی نرمال باشد اما همانطور ی که در گام اول هم پیش بینی کرده بودیم متغیر age نرمال نخواهد بود.

بد نیست نگاهی هم به سطح معناداری بخش Kolmogotov-Smirnov داشته باشیم هرچند این آزمون بیشتر برای مجموعه داده هایی با حجم بالا کاربرد دارد.

گام سوم:

اگر در گام دوم به این نتیجه رسیدید که متغیرهای شما که در گام اول چولگی و کشیدگی معمولی داشتند نرمال نبودند، به سراغ نمودارهایی که می توانید از خروجی Explore که در گام دوم معرفی شد بروید. به طور پیش

فرض نمودارهای جعبه ای ، یا Q-Q plot یا نمودار شاخ و برگ ظاهر می شود. با این حال در همان کادر محاوره Plots چک کنید که گزینه های مربوط به نمودار شاخ و برگ وسایر نمودارها تیک دار باشند.

از طریق این نمودارها می توان به طور جزیی به نحوه ی توزیع داده ها پی برده و علت نرمال نبودن داده ها را با دلیل ببینید:

نمودار شاخ و برگ برای یک توزیع نرمال شکلی زنگوله وار دارد.

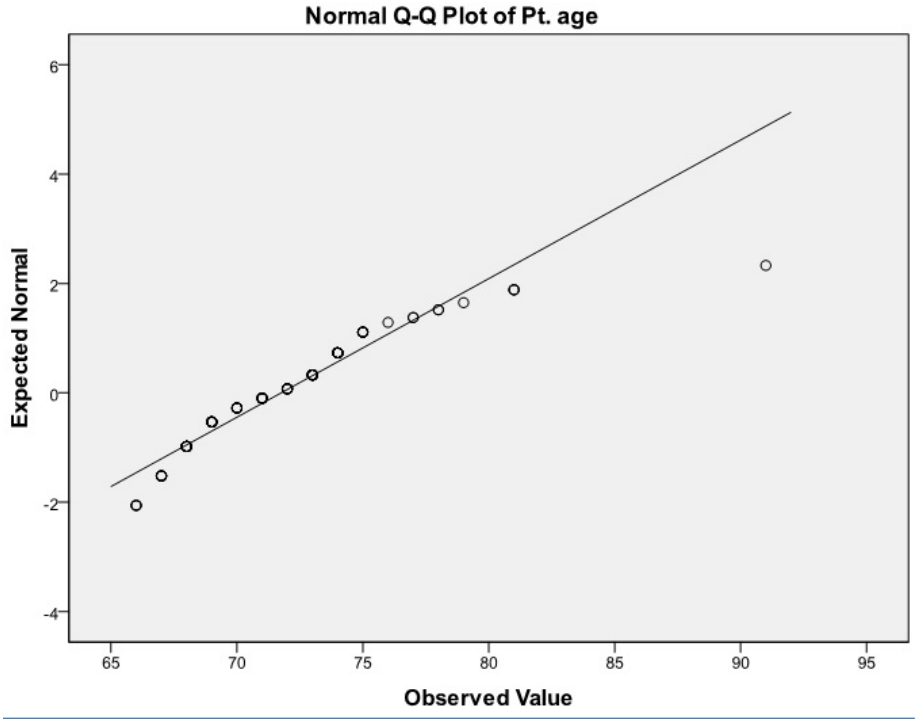
در نمودار Q-Q داده های یک توزیع نرمال معمولا روی خط قرار می گیرد و داده ها در روی خط مارپیچ یا S وار خیلی از نیمساز ربع اول دور نخواهند بود.

نمودار جعبه ی یک نمودار جعبه ای معمولا برای توزیع نرمال در مرکز نمودار قرار می گیرد. دیدن داده های پرت (آنها که با ستاره یا نقطه نشان داد می شوند) می تواند انحراف یک توزیع را از توزیع نرمال به خوبی تشریح کند.

در مثال بالا نمودار ها را برای متغیر age می بینیم:

مشاهده می کنید که علاوه بر اینکه داده ها S – وار حول خط نرمال پخش شده اند داده هایی داریم که بسیار از خط دور افتاده اند. این داده ها همان داده های پرت می باشند

در نمودار جعبه ای آن هم مشاهده می کنید که جعبه در مرکز قرار ندارد. و علت آن می تواند وجود داده ی پرتی که در بالا ی جعبه فرار دارد نیز باشد.



همینطور نمودار شاخ و برگ را برای متغیر سن ببینید و آن را با نمودار شاخ و برگ متغیر los مقایسه کنید:

```
Pt. age Stem-and-Leaf Plot

Frequency Stem & Leaf

3.00 66 . 000
6.00 67 . 000000
14.00 68 . 00000000000000
13.00 69 . 0000000000000
6.00 70 . 000000
8.00 71 . 00000000
6.00 72 . 000000
14.00 73 . 00000000000000
14.00 74 . 00000000000000
6.00 75 . 000000
1.00 76 . 0
2.00 77 . 00
2.00 78 . 00
1.00 79 . 0
.00 80 .
3.00 81 . 000
1.00 Extremes (>=91)

Stem width: 1
Each leaf: 1 case(s)
```

نمودار شاخ و برگ متغیر los:

Hospital LOS Stem-and-Leaf Plot

Frequency Stem & Leaf

3.00 12 . 000
4.00 13 . 0000
12.00 14 . 000000000000
7.00 15 . 0000000
8.00 16 . 00000000
21.00 17 . 000000000000000000000
14.00 18 . 00000000000000
11.00 19 . 00000000000
9.00 20 . 000000000
7.00 21 . 0000000
3.00 22 . 000
.00 23 .
.00 24 .
1.00 25 . 0

Stem width: 1
Each leaf: 1 case(s)

طرح آنالیز واریانس دو طرفه یا طرح دو عاملی

گاهی اوقات یک پژوهشگر به اثر ترکیبی دو درمان علاقه مند میباشد. برای مثال مطالعه ای را در نظر بگیرید که یک پژوهشگر میخواهد میزان کاهش وزن را تحت تاثیر رژیمهای غذایی مختلف و سطوح مختلفی از پیاده روری روزانه بررسی نماید.

	Normal	High Protein	High Fat	Hi Carbohy-drate
0 mi.	8.5 11.5	15.5 16.5	8.5 7.5	15.5 13.5
1 mi.	14 16	20 23	13 11	21 18
2 mi.	24.5 19.5	27 24	22 27	24.5 27.5

۲۴ مردی را که همگی اضافه وزنی در حدود ۴۰ پاوند دارتد را به ۱۲ درمانی که از ترکیب ۴ سطح تغذیه و ۳ سطح پیاده روی بوجود می آید، تخصیص داده میشوند.

در حقیقت این طرح یک طرح اقتصادی میباشد زیرا ما هم زمان اثر دو عامل را (دو روش درمانی یا دو مداخله) یعنی پیاده روی و رژیم غذایی را برروی کاهش وزن افراد بررسی مینماییم. وقتی ما از اقتصادی بودن یک طرح صحبت میکنیم باید مبنایی برای مقایسه داشته باشیم. **یعنی در مقایسه با چه طرحی این روش اقتصادی تر است؟**

ما پیش از این با طرح ANOVA که آنالیز واریانس یک طرفه نیز نامیده میشود آشنا شدیم (در SPSS این طرح با نام One way ANOVA مشخص شده است). اگر ما می خواستیم بر مبنای طرح ANOVA این داده ها را آنالیز نماییم باید این آزمایش را بر اساس دو طرح مجزا تحلیل مینمودیم. یعنی باید ۲ بار ANOVA را به کار می بردیم (یک بار برای بررسی اثر پیاده روی و یک بار برای بررسی اثر رژیم غذایی) و در هر طرح هم باید از نمونه های متفاوتی استفاده می کردیم. این در حالیست که ما در این طرح که آنالیز وارانس دو طرفه نام دارد بر اساس یک نمونه هم زمان اثر این دو عامل را بررسی می نماییم. بنابراین در شرایطی که ما به اثر دو درمان با سطوح مختلف علاقه مند هستیم، استفاده از روش آنالیز وارانس دو طرفه از نظر اقتصادی مفید تر خواهد بود.

اما آیا اقتصادی بودن این طرح تنها دلیل استفاده از آن میباشد؟

جواب این سوال برای یک پژوهشگر هوشمند **(البته نه شما!)** قطعا **نه** خواهد بود. این طرح جدا از اقتصادی بودن آن یک مزیت دیگر نیز دارد و آن اینست که پژوهشگر (باز هم تاکید می نمایم: از نوع آگاه و هوشمند) ممکن است علاوه بر آنکه بخواهد اثر هر یک از دو درمان را در یک طرح و در حضور هم بررسی نماید به اثر متقابل هر دو درمان نیز علاقه مند باشد. بنابراین اقتصادی بودن آنالیز واریانس دو طرفه تنها دلیل استفاده از آن نمیباشد. بعبارتی پژوهشگر ممکن است به این سوال علاقه مند باشد که ترکیب هم زمان این دو درمان چه تاثیری بر کاهش وزن بیماران خواهد داشت. یا به عبارتی روند تغییرات یک درمان در حضور درمان دیگر چگونه خواهد بود (شاید این مفهوم برای اثر متقابل زیباتر باشد). نکته بسیار مهم دیگری که این طرح را از ANOVA متمایز میکند اینست که اگر ما می خواستیم بر مبنای طرح ANOVA این داده ها را آنالیز نماییم باید این آزمایش را بر اساس دو طرح مجزا تحلیل مینمودیم. یعنی باید ۲ بار ANOVA را به کار می بردیم (یک بار برای بررسی اثر پیاده روی و یک بار برای بررسی اثر رژیم غذایی) و در هر طرح هم باید از نمونه های متفاوتی استفاده می کردیم و این امکان وجود داشت که در هر تحلیل ANOVA تفاوت معنی داری بین روشهای تغذیه و یا پیاده روی مشاهده نکنیم. این در

حالیست که در روش آنالیز واریانس دو طرفه چون اثر هر عامل در حضور عامل دیگر سنجیده میشود ممکن است تفاوت معنی داری بین سطوح هر عامل مشاهده نماییم.

در نهایت جمع بندی مطالب بالا به این نکته ختم میشود که آنالیز واریانس دو طرفه به سوالات زیر پاسخ میدهد:

۱- آیا میا نگین کاهش وزن افرادی که از نظر رژیم غذایی متفاوتند، در حضور عامل پیاده روی، از نظر آماری متفاوت است؟

۲- آیا میا نگین کاهش وزن افرادی که از نظر سطح پیاده روی (فعالیت) متفاوتند، در حضور عامل تغذیه، از نظرآماری متفاوت است؟

۳- آیا بین عامل تغذیه و پیاده روی اثر متقابل وجود دارد؟ (برای مثال یعنی وقتی از سطح ۱ رژیم به سطح ۲ رژیم حرکت میکنیم آیا روند تغییرات کاهش وزن در سطوح مختلف فعالیت بدنی (پیاده روی) متفاوت است؟

تحلیل داده ها با SPSS

برای تحلیل این داده ها باید ابتدا داده ها را به روش خاصی در نرم افزار SPSS وارد کنیم. به همین منظور ما ۳ متغیر را در SPSS تعریف مینماییم و داده ها را به شکل زیر در نرم افزار SPSS وارد می کنیم. سپس به قسمت Analyze و بعد به منوی General Liner Model وارد شده و منوی Univariate را انتخاب می کنیم. سپس در قسمت Dependent variable متغیر Response را وارد نموده و در قسمت Fixed Factor(s) متغیر های Diet و Jogging را وارد مینماییم. سپس به قسمت Plots رفته و در قسمت Horizontal axis متغیر Diet و در قسمت Separate Lines متغیر Jogging را وارد میکنیم. سپس بر روی عبارت Add کلیک می کنیم و پس از تایید دستورات فوق خروجی را مشا هده خواهیم کرد.

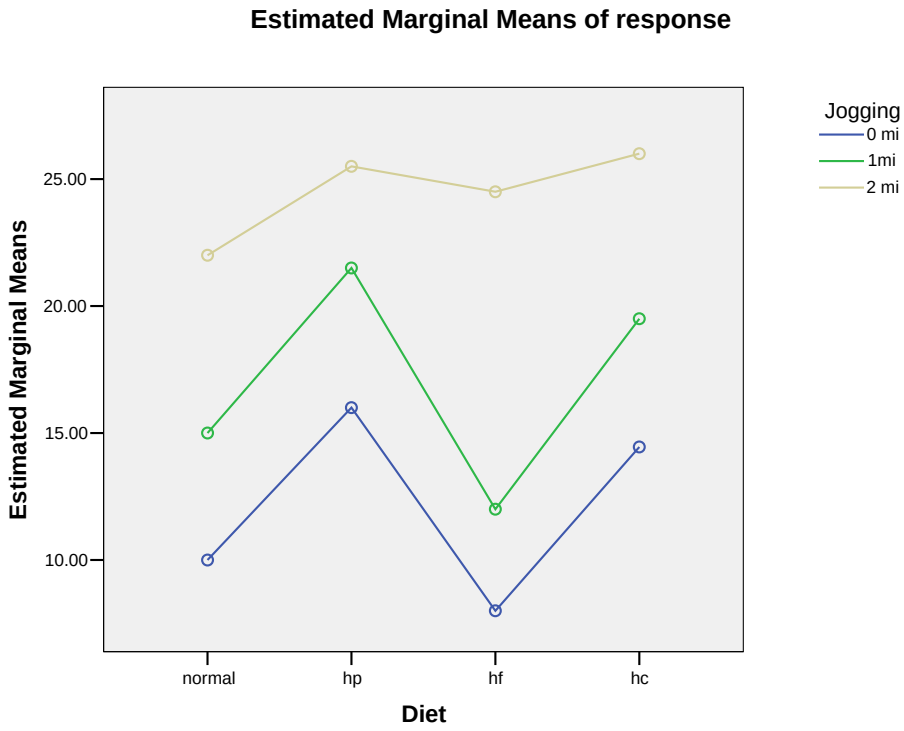
Response	Diet	Jogging
8.50	1.00	.00
11.50	1.00	.00
14.00	1.00	1.00
16.00	1.00	1.00
24.50	1.00	2.00
19.50	1.00	2.00
15.50	2.00	.00
16.50	2.00	.00
20.00	2.00	1.00
23.00	2.00	1.00
27.00	2.00	2.00
24.00	2.00	2.00
8.50	3.00	.00
7.50	3.00	.00
13.00	3.00	1.00
11.00	3.00	1.00
22.00	3.00	2.00
27.00	3.00	2.00
15.40	4.00	.00
13.50	4.00	.00
21.00	4.00	1.00
18.00	4.00	1.00
24.50	4.00	2.00
27.50	4.00	2.00

Tests of Between-Subjects Effects

Dependent Variable: response

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	836.805(a)	11	76.073	16.810	.000
Intercept	7664.800	1	7664.800	1693.723	.000
Diet	170.035	3	56.678	12.524	.001
Jogging	622.901	2	311.450	68.822	.000
Diet * Jogging	43.869	6	7.312	1.616	.226
Error	54.305	12	4.525		
Total	8555.910	24			
Corrected Total	891.110	23			

a R Squared = .939 (Adjusted R Squared = .883)



اگر به مقدار P-value برای Diet و Jogging در جدول ANOVA دقت کنیم مشاهده می کنیم که تفاوت معنی داری بین سطوح رژیم غذایی و پیاده روی وجود دارد. اما اثر متقابل Diet و Jogging معنی دار نیست. در نمودار رسم شده مفهوم اثر متقابل را که معنی دار نیست بهتر شرح میدهد. این نمودار نشان میدهد که وقتی از سطح ۱ رژیم (Normal) به سطح ۲ رژیم (High Protein) حرکت میکنیم روند تغییرات کاهش وزن در سطوح مختلف فعالیت بدنی یکسان است. این مسئله برای حالتی که از سطح ۲ رژیم به سطح ۳ رژیم و همچنین هنگامی که از سطح ۳ رژیم به سطح ۴ رژیم حرکت می کنیم نیز صادق است (به خصوص در سطح 0 mi و 1 mi). که این به مفهوم عدم وجود اثر متقابل بین عامل رژیم غذایی و عامل پیاده روی است.

تحلیل همبستگی

همبستگی(Correlation)

در بسیاری از موارد در انجام مطالعات تحقیقاتی به دنبال بررسی رابطه دو متغیر تصادفی کمُی مـی باشـیم. در ایـن شرایط ارتباط بین دو متغیر کمی (رابطه بین قد و وزن، سن و فشار خون و ...) از شاخصی به نام ضریب همبسـتگی پیرسون استفاده میکنیم. معمولا در بررسی همبستگی دو متغیر مقادیر یکی به دیگری وابسته است. بـرای مثـال در بررسی رابطه سن و فشار خون، فشار خون به سن وابسته است. متغیـر وابسـته را بـا y و متغیـری را کـه مقـادیر آن مقادیر متغیر وابسته را تحت تاثیر قرار میدهد متغیر مستقل مینامیم و با x نمایش میدهیم.

ممکن است در تحلیل همبستگی نتوان هیچ کدام از دو متغیر را به عنوان علت برای دیگری انتخـاب کـرد. در اصـل برای بررسی میزان هماهنگی میان دو متغیر به دنبال شاخصهایی می گردیم که در اصل دو ویژگی: ۱ بـه واحـد دو جامعه وابسته نباشد ۲ کراندار باشد. را داشته باشند:

به طور مثال در تحولات اقصادی به دنبال رابطه میان تقاضای نفت خام در برابر تقاضای طلا می باشیم. یا در مطالعه تحولات اجتماعی به دنبال رابطه درآمد سرپرست خانواده و میزان تحصیل فرزنـدان مـی باشـیم و مثالهـایی از ایـن دست.....

مجموعه اطلاعات (داده های) موجود در انجام یک آزمون همبستگی که شامل اندازه های بدسـت آمـده از دو متغیـر **X** و **Y** می باشند را می توان به صورت یک نمونه تصادفی دو متغیره $(X_1, Y_1), \ldots\ldots\ldots, (X_n, Y_n)$ بیان کرد.

مطالعه رابطه بین متغیرها بوسیله ((تحلیل همبستگی)) (Corroletion Analysis) انجام می شـود. کـه بیـانگر وجود یک رابطه خطی بین دو متغیر می باشد.

فرمول ضریب همبستگی به صورت زیر می باشد

$$r = \frac{Cov(X,Y)}{S_x S_y} = \frac{\sum_{i=i}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2\right]\left[\sum_{i=1}^{n} (y_i - \overline{y})^2\right]}}$$

ضریب همبستگی همواره مقداری بین صفر و یک دارد: $-1 \le r \le +1$.

با توجه به مقدار r در حالتهای مختلف تفسیرهای گوناگونی از رابطه X و Y خواهیم داشت.

حالتهای مختلف برای r :

۱ $r = 1$ در این حالـت همبسـتگی کامـل و مسـتقیم گـوییم. بـا افـزایش مقـدار x مقـدار y بـه طـور قطعـی زیاد می شود.

۲ $r = -1$ در این حالت همبستگی را کامل و معکوس گوییم. با افزایش مقدار x مقدار y کاهش می یابد.

۳ $-1 < r < 0$ همبستگی ناقص و معکوس است. با افزایش مقدار x مقدار y به طور نسبی کاهش می یابد.

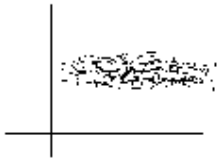
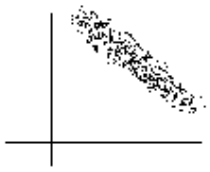
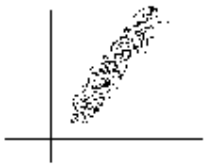
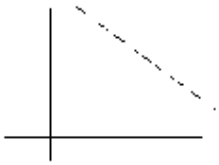
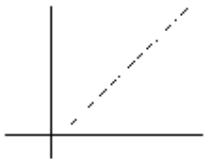
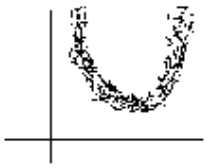
۴ $0 < r < 1$ همبستگی ناقص و مستقیم است. با افزایش مقدار x مقدار y به طور نسبی افزایش می یابد.

۵ $r = 0$ ۱ رابطه خطی وجود ندارد. (به طور مثال رابطه ممکن است از نوع درجه دو باشد)

۲ شیب خط صفر می باشد.

توجه کنید که صفر بودن ضریب همبستگی به معنی عدم ارتباط بین دو متغیر نیست بلکه به مفهوم عدم ارتباط خطی بین دو متغیر است. دلیل این امر نیز آنست که فرمول ضریب همبستگی پیرسون تنها رابطه خطی را اندازه گیری میکند.

اولین گام در بررسی رابطه دو متغیر کمی رسم نمودار پراکندگی داده ها (Scatter Plot) میباشد. این نمودار میتواند به شکلهای مختلف مشاهده شود.

الف **r = 0** 	ب **r = 0** 
ج **r<0** 	د **r>0** 
ه **r = -1** 	و **r=1** 
ز در این حالت نمیگوییم ارتباطی بین x وy وجود ندارد. بلکه رابطه خطی بین وجود ندارد **r = 0** 	

روش ورود اطلاعات در نرم افزار SPSS

در این روش باید اطلاعات قبل و بعد در دو ستون متفاوت و در مقابل هم همانند آنچه که در جدول بالا مشاهده شد در نرم افزار SPSS وارد شوند. سپس به قسمت Analyze در منوی های SPSS رفته و قسمت Correlate را انتخاب نموده و در این قسمت بر روی عبارت Bivariate کلیک نمایید و دو متغیر موجود در سمت راست را به قسمت variables در سمت چپ منتقل نموده و دستور OK را اجرا نمایید.

انواع ضریب همبستگی با توجه به نوع متغیرهای مورد مطالعه

۱- پیرسن : در این روش متغیر X و Y هر دو پیوسته می باشند. (در پنجره ***Variable View*** در ستون مربوط به مقیاسها(***Measure***) داده ها باید از نوع ***Scale*** انتخاب شوند.)

۲ کندال : در این روش هر دو متغیر **X** و **Y** باید به صورت طبقه بنـدی شـده باشـد یعنـی(***Ordinal,Nominal***)
(در پنجره Variable Viewدر ستون مربوطه باید داده ها از نوع Ordinal انتخاب شوند.)

۳ اسپیرمن : در این روش متغیر **X** گسسته و متغیر **Y**پیوسته می باشد.
(متغیر X در پنجـره ***Variable View***در سـتون مربوطـه بایـد از نـوع ***Ordinal*** و متغیـر Y از نـوع ***Scale*** انتخاب شوند.)

در آزمونهای بالا متغیر **X** را به عنوان متغیر مستقل و متغیر **Y** را به عنوان متغیر وابسته در نظر می گیریم.

- **نکته:. در بعضی مواقع می توانیم با طبقه بندی داده های پیوسته آنها را به صورت طبقه بندی شده در آوریم.**

به طور مثال اگر بخواهیم تأثیر درآمد خانواده را بر روی معدل فرزندان بررسی کنیم به دلیل اینکـه درآمـد و معـدل هر دو متغیر پیوسته می باشند باید از ضریی همبستگی پیرسن استفاده کنیم. اما می توانیم با تقسیم بنـدی درآمـد به طور مثال به سه گروه کم درآمد، متوسط و پر درآمد از ضریب همبستگی اسپیرمن استفاده کنیم.

در آزمونهای همبستگی یک طرفه فرضهای زیر بررسی می شوند.

$$1\begin{cases} H_o: \ r = o \\ H_1: \ r > o \end{cases} \qquad 2\begin{cases} H_o: \ r = o \\ H_1: \ r < o \end{cases}$$

جهت نامساوی با توجه به برآورد نمونه ای r تعیین می شوند. اگـر r مثبـت باشـد آزمـون (۱) و اگـر منفـی باشـد از آزمون (۲) استفاده می کنیم.

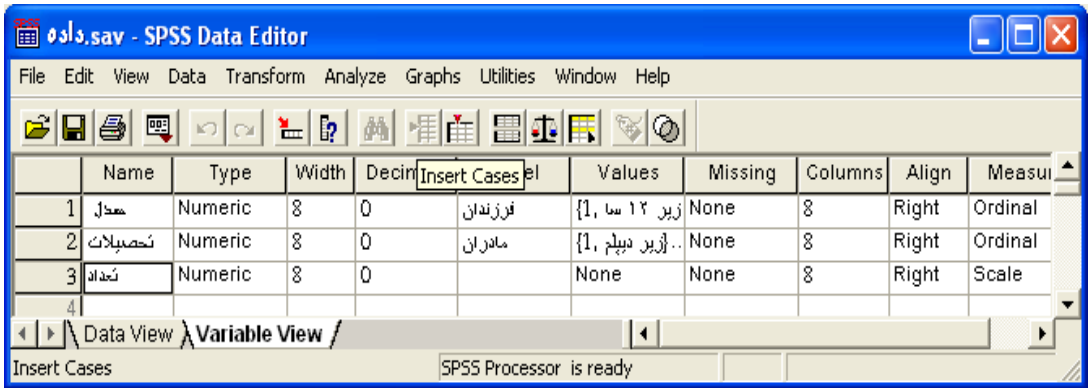
مقدار پیرسون بیانگر شدت رابطه است. یعنی اگر مقدار پیرسون بین صفر تا ۳/ باشد ضعیف و اگر بین ۳/ تا ۶/ باشـد متوسط و اگر بیشتر از ۶/ باشد قوی است. این معیار قراردادی است ولی معمولا در تفسیر نتایج و تحقیقـات مختلـف از آن استفاده می گردد.

مثال: در جدول زیر میزان معدل دانش آموزان و میزان تحصیلات آنها آمده است. هدف تعیین میـزان همبسـتگی و نوع ارتباط معدل با میزان تحصیلات مادر می باشد

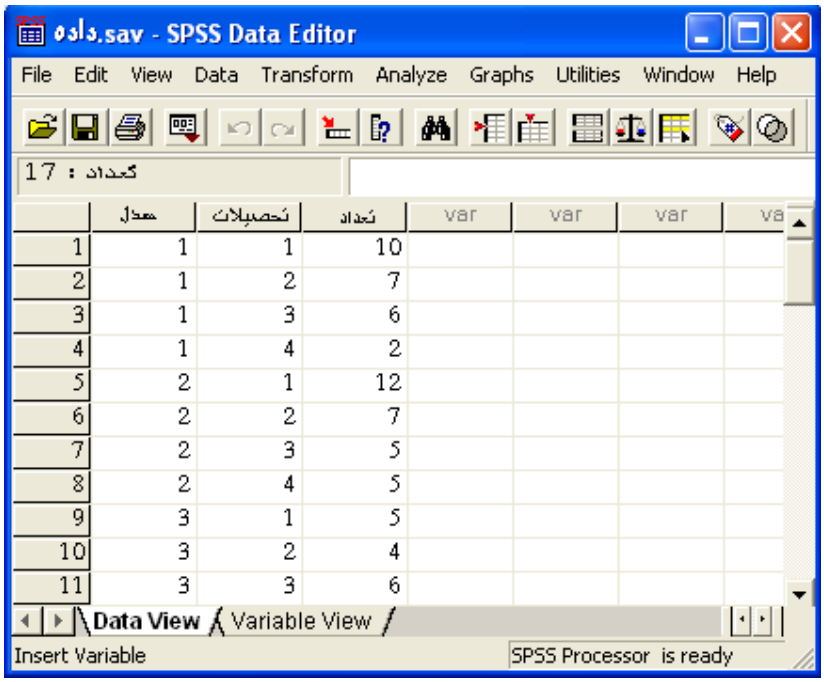
.

تحصیلات / معدل	(۱) زیردیپلم	(۲) دیپلم	(۳) فوق دیپلم	(۴) لیسانس وبالاتر
(۱) زیر ۱۲	۱۰	۷	۶	۲
(۲) ۱۲ تا ۱۵	۱۲	۷	۵	۵
(۳) ۱۵ تا ۱۷	۵	۴	۶	۱۰
(۴) ۱۷ تا ۲۰	۵	۷	۱۰	۱۲

وقتی در نرم افزار Spss داده ها را به صورت رتبه ای مشخص کردیم می توانیم در تجزیه و تحلیلها تفسیر راحتـری از خروجیها داشته باشیم.

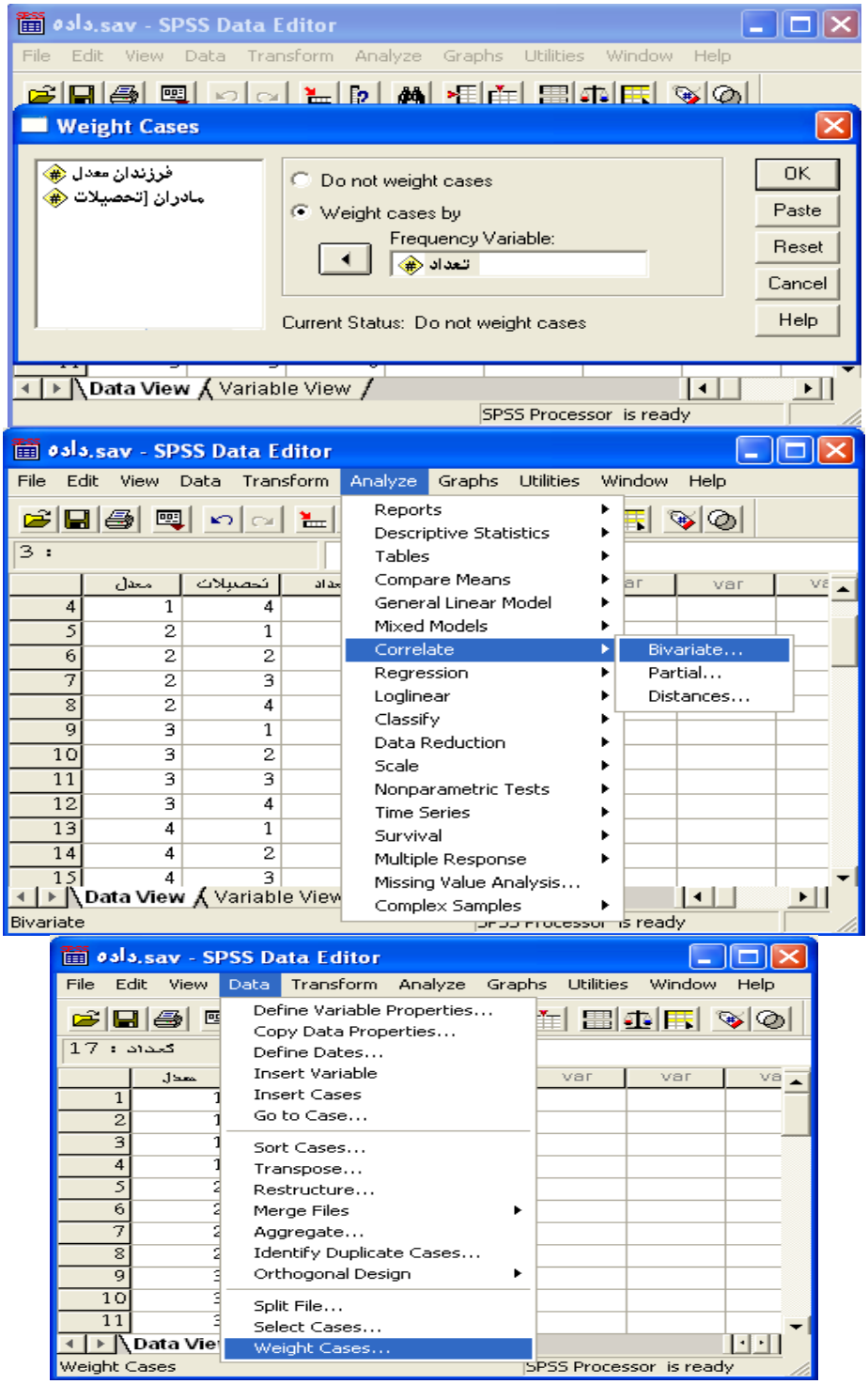


به مانند مثال قبلی در پنجره ***Variable View*** دو متغیر معدل و تحصیلات را وارد می کنیم و ستونها را متناسب با نوع متغیرها تنظیم می کنیم. در ستون ***Values*** با توجه به کدهایی که در جدول بالا داده شده است متغیرها را معرفی می کنیم.

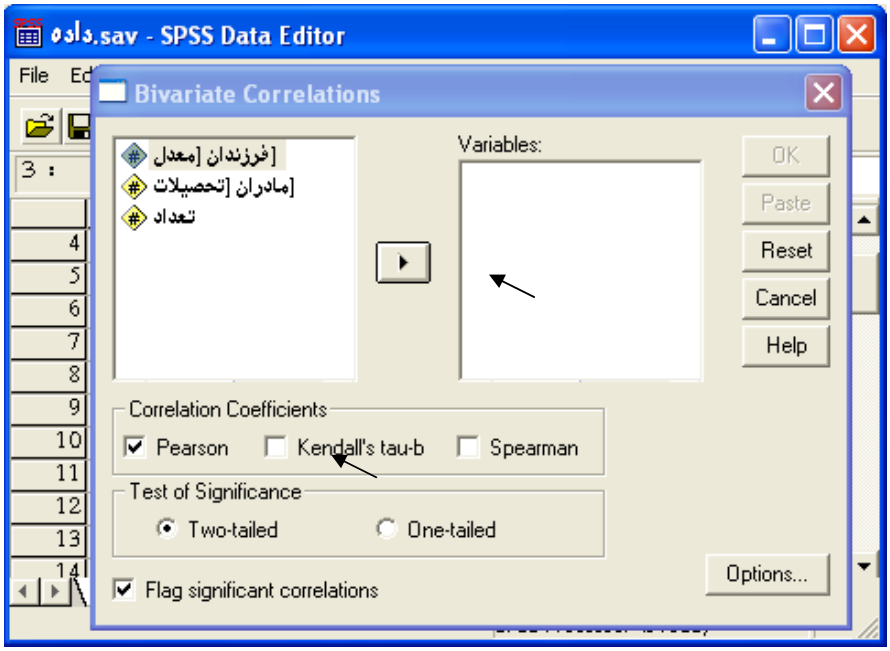


سپس داده را به صورت زیر در پنجره ***Data View*** وارد می کنیم

برای اینکه مشخص کنیم ستون تعداد معرف فراوانی هایی هست که به دو متغیر سن و تحصیلات نسبت داده شده است و و از نوع اندازه گیری شده نمی باشند باید دو ستون تحصیلات و سن توسط ستون تعداد وزن دار شود. برای این کار به صورت زیر عمل می کنیم.



با کلیک بر روی کلمه Ok داده ها بوسیله ستون تعداد وزند دار می شوند. سپس مسیر زیر را انتخاب می کنیم Bivariate Correlation باز شود.



با بردن دو متغیر سن فرزندان و تحصیلات مادران به مربع سمت راستی (***Variable***) و انتخاب ضریب همبستگی متناسب با داده ها، در این سوال (فعال کردن گزینه ***Kendall's tau-b*** در قسمت ***Correlation Coefficients***) و انتخاب نوع آزمون (یک طرفه بودن(یک دامنه) یا دو طرفه بودن(دو دامنه) در قسمت ***Test of significance***) با کلیک کردن بر روی ***Ok*** خروجیهای مربوط به آزمون را به صورت زیر مشاهده کرد.

اگر فرضیه شما دو دامنه است یعنی جهتی در فرضیه مشخص نکرده اید در تیک های بالا، دودامنه(tailed ۲) که بصورت پیش فرض فعال است را مارک دار انتخاب کنید.

مقدار ***P-Values*** بدست آمده برابر است ۰/۰۰۱ می باشد که با توجه به p مقدار(***2-tailed***).sig با ۰/۰۵ مقایسه می شود و چون $P<0.05$ می باشد نتیجه می شود که یک نوع رابطه بین معدل دانش آموزان و تحصیلات مادران آنها وجود دارد و چون $r = 0.260$ بدست آمده است و $0<r<1$ می باشد نتیجه می گیریم که همبستگی از نوع مستقیم و ناقص می باشد و اینگونه تفسیر می شود که با افزایش تحصیلات مادران (با رفتن از کد ۱ به سمت کد ۴) معدل دانش آموزان نیز (با رفتن از کد ۱ به سمت کد ۴) افزایش پیدا می کند و نشان دهنده نقش موثر و مثبت تحصیل مادر بر بالا بردن کیفیت تحصیلی فرزندانشان می باشد. در این مثال اگر تک تک معدلها را به همراه نوع مدرک مادر در اختیار داشتیم، در یک ستون نوع مدرک و در ستون دیگر داده های پیوسته معدل را وارد می کردیم و برای انجام آزمون از ضریب همبستگی اسپیرمن استفاده می کردیم.

انتخاب مناسب آزمون آماری در spss

یکی از مهم ترین سوالاتی که در انجام پژوهش های آماری مطرح می شود انتخاب مناسب روش تحلیل داده ها است. معمولا نوع آزمون های آماری از فاکتورهای ویژگیهای نمونه، انواع داده ها و همچنین نوع تحلیلی که مورد نیاز آمارگر است، مشخص می شود. این تحلیل ها به دو بخش عمده ی آمار ناپارامتری و پارامتری بر می گردد. در تحقیقات میدانی، حجم زیادی از داده های آماری به عنوان نمونه گردآوری می شود که از طریق آنها می توان به سوالات و فرضیه های تحقیق پاسخ داد. سوالات تحقیق بایستی براساس اهداف تحقیق (هدف کلی و جزیی) طراحی شوند تا نتیجه ی تجزیه و تحلیل آنها، بیانگر واقعی هدف تحقیق باشد. معمولا اگر محقق بدون به کار بردن آمار به تحلیل داده ها بپردازد نتیجه آن واقعی نخواهد بود و محقق به مشکل برخواهد خورد. لذا دانستن آمار و نکات ریز آماری از ضروریات فرایند تحقیق و پژوهش است .

استنباط آماری در واقع یک نوع نتیجه گیری کلی از جز به کل است و با آزمایش و خطا همراه است. یک جنبه از استنباط آماری محاسبه برآوردهایی از پارامترهای جامعه است مثل میانگین یا واریانس جامعه از طریق آماره های نمونه مانند میانگین یا واریانس نمونه.

در آمار استنباطی روش های گوناگونی برای تجزیه و تحلیل داده های آماری وجود دارد که هر یک برای نوع خاصی از داده ها و همین طور نوع خاصی از سوالات مورد استفاده قرار می گیرند .

انتخاب یک آزمون آماری را وابسته به موارد زیر می دانند:

۱ نقشه یا طرح تحقیق

۲ هدف تحقیق

۳ طبیعت داده ها (تعداد متغیرها، کمی یا کیفی، پیوسته یا گسسته)

۴ توزیع داده ها (نرمال بودن یا نبودن متغیرهای مورد بررسی)

۵ نوع استنباط مورد نیاز (توصیف جامعه، مقایسه اختلاف دو یا چند گروه، سنجش رابطه بین دو یا چند متغیر)

به طور کلی هدف هر آزمون آماری به دنبال این هدف هستیم که آیا داده های نمونه، دلایل کافی برای رد یک فرضیه در مورد جامعه ارائه می دهد یا خیر؟

بهتر است قبل از انتخاب یک آزمون آماری بتوانیم به سوالات زیر پاسخ دهیم:

فرضیه تحقیق چیست؟

نوع داده ها چیست؟

توزیع ویژگی مورد بررسی در جامعه نرمال است یا خیر؟

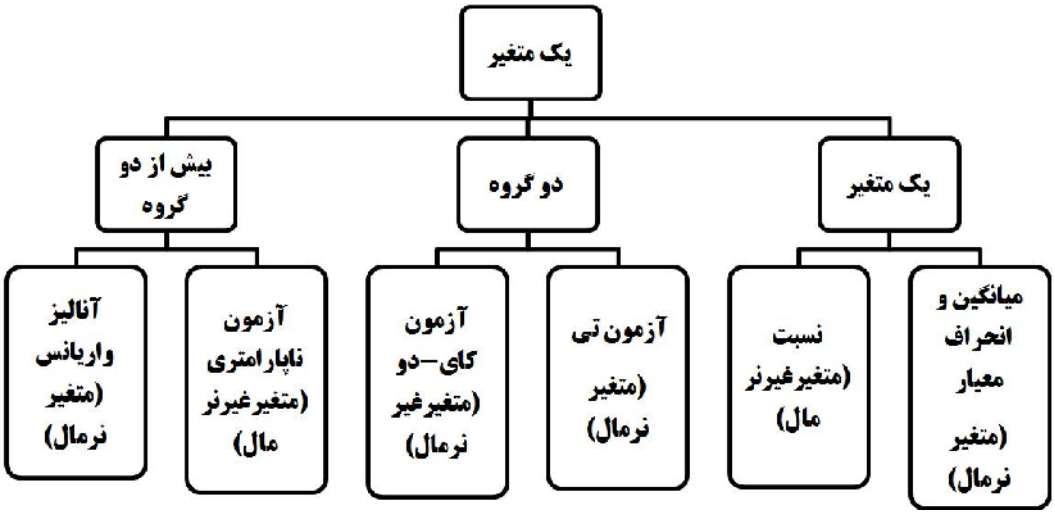
چه تعداد متغیر بررسی می شوند؟

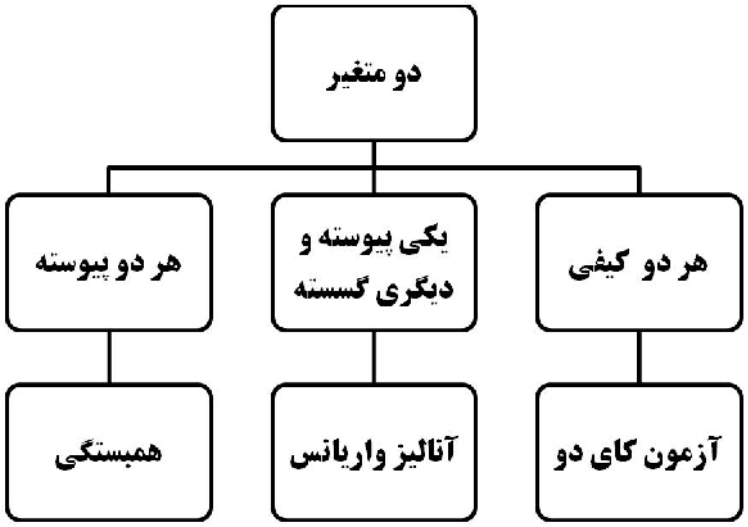
چند گروه مقایسه می شوند؟

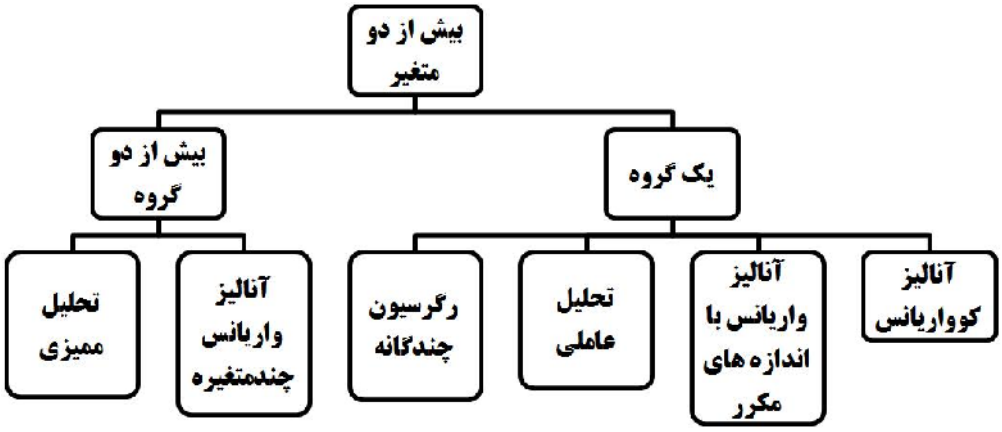
گروه های مورد بررسی مستقلند یا خیر؟

در حالت کلی قبل از هر چیز باید به نرمالیتی توزیع متغیر بپردازیم؛ به طور کلی اگر توزیع متغیر نرمال باشد از آزمون های t و یا آنالیز واریانس و در صورت نرمال نبودن توزیع متغیر از آزمون های ناپارانتری استفاده می شود .

در نمودارهای زیر نحوه ی انتخاب آزمون ها را به ترتیب برای یک و دو و بیشتر از دو متغیر را رسم کرده ایم.


یک متغیر
یک متغیر
دو گروه
بیش از دو گروه
میانگین و انحراف معیار (متغیر نرمال)
نسبت (متغیرغیرنر مال)
آزمون تی (متغیر نرمال)
آزمون کای–دو (متغیرغیر نرمال)
آزمون ناپارامتری (متغیرغیرنر مال)
آنالیز واریانس (متغیر نرمال)


دو متغیر
هر دو کیفی
یکی پیوسته و دیگری گسسته
هر دو پیوسته
آزمون کای دو
آنالیز واریانس
همبستگی


بیش از دو متغیر
یک گروه
بیش از دو گروه
آنالیز کوواریانس
آنالیز واریانس با اندازه های مکرر
تحلیل عاملی
رگرسیون چندگانه
آنالیز واریانس چندمتغیره
تحلیل ممیزی

باید توجه کرد زمانی که دو یا بیش از دو متغیر مد نظر تحلیل است علاوه بر ویژگیهای خود داده ها؛ رابطه ی میان هر دو متغیر نیز مهم است. اینکه جنس دو متغیر چه باشد؟ میان آنها همبستگی وجود دارد یا خیر؟ سوالاتی هستند که باید از خود قبل از انتخاب آزمون ها پرسید. همانطور که در نمودار ۲ نشان دادیم، برای بررسی رابطه ی بین دو متغیر وقتی هر دو متغیر پیوسته باشند از آزمون ضریب همبستگی برای بررسی رابطه بین آنها استفاده می کنیم و چنانچه یکی از متغیرها پیوسته و دیگری گسسته باشد از آنالیز واریانس و در حالتی که هر دو متغیر کیفی باشند از آزمون کای دو کمک می گیریم. همینطور در نمودار ۳ به راحتی می توان دریافت در صورتی که متغیرها برای یک گروه به کار برده شوند از تحلیل عاملی یا آنالیز واریانس با اندازه های مکرر و یا رگرسیون چندگانه استفاده می شود. از تحلیل ممیزی و آنالیز واریانس چندگانه نیز برای بررسی بیش از دو گروه استفاده می شود.

در اینجا به بسیاری از سوالاتی که مطرح می شوند پاسخ می دهیم:

بسیاری از آزمون های آماری بر فرض نرمالیتی جامعه استوارند آزمون هایی که با این جامعه ها سر و کار دارند آزمون های پارامتری هستند مانند آزمون t ، و در مقابل آزمون های ناپارامتری دیگر فرض نرمالیتی را ندارند مانند آزمون های رتبه ای از قبیل آزمون من ویتنی، یا کروسکال والیس.

برخی مواقع به راحتی می توان بر اساس نرمال بودن یا نبودن توزیع متغیرها گفت که کدام نوع از آزمون های پارامتری یا ناپارامتری را به کار گیریم. اگر متغیر مورد بررسی رتبه ای باشد مثل رتبه بندی رضایت کارمندان(کم متوسط زیاد) آن گاه توزیع نرمال ندارد لذا از آزمونهای ناپارامتری بهره می گیریم.

اما نمی توان گفت که وقتی داده ها کمی هم باشند حتما از آزمونهای پارامتری استفاده می کنیم. در مواردی پیش می اید که داده ها کمی بوده اما تعدادشان اندک است در این مواقع، آزمون کولموگروف اسمیرنف هم نمی تواند نرمال بودن داده ها را تعیین کند چرا که برای داده هایی با حجم بزرگتر کاراست. لذا تشخیص اینکه این داده ها نرمال هستند یا خیر بسیار مشکل است در این مواقع برخی از آماردانان از آزمون های پارامتری استفاده کرده و ادعا دارند که در نرمال بودن داده ها مشکلی ایجاد نمی شود. اما برخی دیگر آزمون های ناپارامتری را به کار می برند.

جدول زیر می تواند شما را در انتخاب آزمون مناسب استقلال یاری کند...

	دو گروه		بیش از دو گروه	
متغیر	مستقل	وابسته	مستقل	وابسته
کمی	آزمون t مستقل	آزمون t زوجی	آنالیز واریانس یکطرفه	واریانس در تکرار مشاهدات
رتبه ای	من–ویتنی	ویلکاکسون– آزمون علامتی	کروسکال والیس	فرید من
اسمی	کای دو	مک نمار	کای دو	کوکران

جدول زیرنیز شما را در انتخاب آزمون های آماری مناسب در همبستگی آنها یاری می کند.نوع ضریب همبستگی ؛ بسته به نوع متغیر ها تعیین می شود.

نوع متغیر اول	نوع متغیر دوم (پیشگو)	نوع آزمون
کمی	کمی	ضریب همبستگی پیرسون
کمی /رتبه ای	رتبه ای/کمی	ضریب همبستگی اسپیرمن–کندال
رتبه ای	رتبه ای	ضریب همبستگی اسپیرمن–کندال
کمی	اسمی	ضریب اتا
اسمی	کمی	ضریب لاندا
رتبه ای	اسمی	ضریب تتا
اسمی	اسمی	ضریب لاندا، فی و کرامر